

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2014 年 4 月 3 日(03.04.2014)



(10) 国際公開番号
WO 2014/049995 A1

- (51) 国際特許分類:
G06F 21/60 (2013.01)
- (21) 国際出願番号: PCT/JP2013/005392
- (22) 国際出願日: 2013 年 9 月 12 日(12.09.2013)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2012-212454 2012 年 9 月 26 日(26.09.2012) JP
- (71) 出願人: 日本電気株式会社(NEC CORPORATION)
[JP/JP]; 〒1088001 東京都港区芝五丁目 7 番 1 号
Tokyo (JP).
- (72) 発明者: 高橋 翼(TAKAHASHI, Tsubasa); 〒
1088001 東京都港区芝五丁目 7 番 1 号日本電気
株式会社内 Tokyo (JP).
- (74) 代理人: 下坂 直樹(SHIMOSAKA, Naoki); 〒
1088001 東京都港区芝五丁目 7 番 1 号日本電気
株式会社内 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,

BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS,
LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,
PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,
SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, ZA, ZM, ZW.

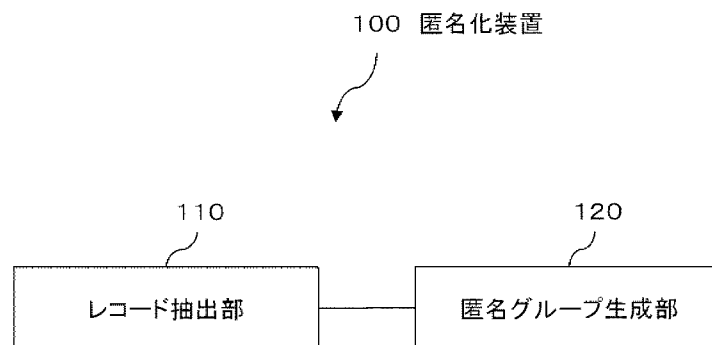
- (84) 指定国 (表示のない限り、全ての種類の広域保
護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW,
MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシ
ア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ
(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,
GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT,
NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML,
MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

(54) Title: INFORMATION PROCESSING DEVICE THAT PERFORMS ANONYMIZATION, ANONYMIZATION METHOD,
AND RECORDING MEDIUM STORING PROGRAM

(54) 発明の名称: 匿名化を実行する情報処理装置、匿名化方法及びプログラムを記録した記録媒体



100 Anonymization device
110 Record extraction unit
120 Anonymous group generation unit

(57) Abstract: The present invention provides an information processing device that performs anonymization such that information about the correspondence relationships between records does not become too unclear. This information processing device is equipped with: a means that extracts multiple second record combinations from combinations of a first record containing a first attribute and a second record containing a second attribute, with the first and second records having the same unique identifier, said extraction being performed on the basis of the ability to satisfy a respective first and second I-diversity in each first record group corresponding to a second record group, and on the basis of the level of abstraction of the correspondence relationships existing between the first and the second records; and a means that generates a data set of an anonymous group comprising a combination of those second records, so as to satisfy the second I-diversity in the combination of those second records and so as to satisfy the first I-diversity in a combination of the corresponding first records.

(57) 要約:

[続葉有]



WO 2014/049995 A1

本発明は、レコード間の対応関係の情報が曖昧になりすぎないように匿名化を実行する情報処理装置を提供する。その情報処理装置は、固有識別子を同一として対応する第1の属性を含む第1のレコードと第2の属性を含む第2のレコードとの組の中から、第2のレコード群に対応する第1のレコード群のそれぞれにおいて第1及び第2のそれぞれの1-多様性を充足可能であることと、第1及び第2のレコードとの間に存在する対応関係の抽象度とに基づいて、複数の第2のレコードの組を抽出する手段と、その第2のレコードの組からなる匿名グループのデータセットをその第2のレコードの組において第2の1-多様性を充足し、対応する第1のレコードの組において第1の1-多様性を充足するように、生成する手段を備える。

明 細 書

発明の名称： 匿名化を実行する情報処理装置、匿名化方法
及びプログラムを記録した記録媒体
技術分野

[0001] 本発明は、個人情報等のように、オリジナルな情報内容のままで公開や利用されることが好ましくない情報を匿名化する情報処理装置、匿名化方法、及びそのためのプログラムに関する。

背景技術

[0002] 購買履歴や診療履歴等のように、サービス提供者によって日々ユーザ（利用者）に提供されるサービス活動から生まれるログ情報は、それらのサービス提供者によって、履歴情報として蓄積されている。これらの履歴情報を分析することで、特定の利用者の行動パターンの把握、ある集団が持つ固有の傾向の把握、将来起こり得る事象の予測、及び過去の事象に対する要因分析等が可能である。これらの履歴情報及びその分析結果を利用することで、サービス提供者は、自己の事業の強化や見直しが可能である。よって、履歴情報は、利用価値が非常に高い有益な情報である。ここで、ある集団は、複数の利用者からなる集団である。

[0003] サービス提供者以外の第三者にとっても、このようなサービス提供者が保有する履歴情報は有益である。例えば、その第三者は、係る履歴情報を利用することで、自身では得られなかった情報を手にすることができる。従って、この第三者自身のサービスやマーケティングの強化が可能である。また、サービス提供者自身が、その第三者に対して、履歴情報の分析を依頼する場合や、研究を目的として、係る履歴情報を公開する場合もある。

[0004] 上述のように利用価値の高い履歴情報には、その履歴情報の主体にとって他人に知られたくない情報や、第三者に知られるべきでない情報が含まれている場合がある。このような情報は、一般に、センシティブ情報（機微情報：Sensitive Attribute (SA)、Sensitive Value）と呼ばれる。例えば、購買履歴の場合は、購入した商品が

センシティブ情報に成り得る。診療情報の場合は、傷病名や、診療行為名がセンシティブ情報である。

[0005] 履歴情報には、サービス利用者を一意に識別するユーザ識別子（ユーザID）と、サービス利用者を特徴付ける複数の属性（属性情報）とが付与されている場合が多い。ユーザ識別子には、氏名、会員番号や被保険者番号などが該当する。サービス利用者を特徴付ける属性には、性別、生年月日、職業、居住エリア、郵便番号などが該当する。サービス提供者は、これらのユーザ識別子と、複数種類の属性と、センシティブ情報とを、一つのレコードとして記録する。そして、サービス提供者は、係るレコードを、該当するユーザ（サービス利用者）がサービスを享受する度に、履歴情報として蓄積する。ユーザ識別子が付与されたままの履歴情報が第三者に提供されると、その第三者はそのユーザ識別子を用いることによってサービス利用者を特定することが可能である。このため、プライバシー侵害の問題が、発生し得る。

[0006] また、複数のレコードによって構成されるデータセットの中から、各レコードに付与されている属性値を1つ以上組み合わせることにより、ある個人を特定できてしまう場合がある。このように個人を特定し得る属性は、準識別子（*Quasi-Identifier*）と呼ばれる。即ち、例えユーザ識別子を取り除いた履歴情報であっても、準識別子に基づいてある個人を特定可能であれば、プライバシー侵害が発生し得る。

[0007] 但し、その一方で、全ての準識別子を履歴情報から取り除いてしまうと、統計的な分析が不可能になる。従って、その履歴情報の本来の有益性が大幅に失われる。例えば、その統計的な分析とは、その全ての準識別子取り除かれた履歴情報に対する分析である。具体的には、ある年代が好んで購入する傾向にある製品の分析や、ある地域に居住する住民が罹患する特有の傷病の分析等が行えない。

[0008] このような特性を有する履歴情報のデータセットを、本来の有用性を保ちながら、プライバシーを保護した形態に変換する手法として、匿名化（匿名化技術：*Anonymization*）が知られている。

- [0009] 例えば、特許文献1は、入力されたデータを属性毎に準識別子或いは重要情報に分類し、全てのその準識別子における“ k -匿名性”と全てのその重要情報における“ l -多様性”を満たすデータセットを出力する技術を開示する。
- [0010] 非特許文献1は、最もよく知られた匿名性指標である k -匿名性を提案する。匿名化対象のデータセットに、係る k -匿名性を充足させる手法は、“ k -匿名化”と呼ばれる。この k -匿名化では、同じ準識別子を有するレコードがその匿名化対象のデータセットの中に少なくとも k 個以上存在するように、対象となる準識別子を変換する処理が行われる。この変換処理としては、一般化、切り落とし等の方式が知られている。係る一般化において、元の詳細な情報は、抽象化された情報に変換される。
- [0011] 非特許文献2は、係る k -匿名性を発展させた匿名性指標の1つである、 l -多様性を提案する。匿名化対象のデータセットに、係る l -多様性を充足させる手法は、“ l -多様化”と呼ばれる。この l -多様化では、同じ準識別子を持つ複数のレコードに、少なくとも l 種類以上の異なるセンシティブ情報が含まれるように、対象となる準識別子を変換する処理が行われる。
- [0012] ここで、 k -匿名化は、準識別子と関連付けされるレコードの数が k 個以上になることを保証する。また、 l -多様化は、準識別子と関連付けされるセンシティブ情報の種類が l 種類以上になることを保証する。
- [0013] 上述した k -匿名化や、 l -多様化では、同一のユーザ識別子を持つ複数のレコードが存在する場合に、それらレコード間の順序や関係等の、互いに異なる事象間の対応関係（換言すれば、特徴、遷移、プロパティ：以下、本願では「対応関係」と称する）が考慮されていない。そのため、係るレコード間の性質が曖昧になったり、失われたりしてしまう場合がある。
- [0014] また、同一のユーザ識別子を持つ複数のレコードを対象とした、時間軸上における順序を保存した匿名化方法として、移動軌跡に対する匿名化技術が知られている。
- [0015] 非特許文献3は、位置情報が時系列に関連付けされた移動軌跡を匿名化す

る技術に関する論文である。より具体的に、非特許文献3に記載された匿名化技術は、係る移動軌跡の始点から終点までを一連のシーケンスとみなして、一貫した k -匿名性を保証する匿名化技術である。この移動軌跡の匿名化技術では、地理的に類似する k 個以上の移動軌跡を束ねたチューブ状の匿名移動軌跡が生成される。移動軌跡の匿名化技術では、匿名性の制約の中で、地理的な類似性を最大化した匿名移動軌跡が生成される。

[0016] 非特許文献3に代表される移動軌跡の匿名化方式では、同一のユーザ識別子を与えられたレコード間に存在する性質のうち、特に、時系列な順序関係が保たれる。

先行技術文献

特許文献

[0017] 特許文献1：特開2012-003440号公報

非特許文献

[0018] 非特許文献1：L. Sweeney, “ k -anonymity: a model for protecting privacy”, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), pp. 555-570, 2002.

非特許文献2：A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam, “ l -Diversity: Privacy Beyond k -Anonymity”, ACM Transactions on Knowledge Discovery from Data, Volume 1 Issue 1, March 2007 Article No. 3.

非特許文献3：O. Abul, F. Bonchi and M. Nanni, “Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases.” In Proceedings of 24th IEEE

E International Conference on Data Engineering, pp. 376-385, 2008.

発明の概要

発明が解決しようとする課題

- [0019] しかしながら、上述した特許文献及び非特許文献に記載された技術においては、対応関係の情報を含むデータセットが１ー多様性を充足するように匿名化を施された場合、その情報が曖昧になりすぎる場合があるという問題点がある。ここで、「対応関係」の情報とは、「同一の固有識別子（ユーザ識別子）を有するレコード間の対応関係」の情報である。ここで、そのデータセットは、例えば、同一の固有識別子をそれぞれが有するレコードの組を、１以上含む、複数のレコードから構成されるデータセットである。
- [0020] そのデータセットにおいて、例えば、それらのレコードの一部からなるレコード群毎に１ー多様性が定められる。そして、それらの１ー多様性を充足するようにデータセットが匿名化される。そうした場合に、その匿名化されたデータセットに含まれる「同一の固有識別子を有するレコード間の対応関係」が、元のデータセットのそれに比べて曖昧になりすぎる場合がある。
- [0021] その情報（「対応関係」の情報）が曖昧になりすぎる場合がある理由は、以下のとおりである。
- [0022] 上述の特許文献及び非特許文献に記載の技術は、「同一の固有識別子を有するレコード間の対応関係」の情報を、維持するために必要な考慮を払っていない。そのため、それらのレコード群毎に定められた１ー多様性を充足するように、データセットが匿名化された場合、元のデータセットには存在しなかった、余分な「同一の固有識別子を有するレコード間の対応関係」が付加される場合がある。
- [0023] 特許文献１では、「同一の固有識別子を有するレコード間の対応関係」の情報を考慮していない。
- [0024] 非特許文献１は、１ー多様性に関する技術を開示していない。
- [0025] 非特許文献２では、地理的な類似性を最大化した匿名移動軌跡を構築する

ことを主たる目的としており、必ずしも、各レコード間の性質（対応関係）が維持されるわけではない。また、非特許文献3では、1ー多様性の匿名性の保証には対応していない。

[0026] 次に、具体的な例を説明する。

[0027] 図28は、匿名化前データセットの一例を示す図である。図28に示す匿名化前データセットは、複数の第1のレコードと複数の第2のレコードとを含む。その第1のレコードは、固有識別子と診療月、年齢及び病名の属性とを含み、診療月の属性値が「4月」である。その第2のレコードは、固有識別子と診療月、年齢及び病名の属性とを含み、診療月の属性値が「5月」である。

[0028] また、図28に示す匿名化前データセットは、同一の固有識別子を有する第1のレコードと第2のレコードとの間の対応関係の、情報を含んでいる。例えば、その対応関係は、固有識別子が「1」の第1のレコード及び第2のレコードのそれぞれに含まれる病名の属性値の「U」と「A」との対応関係（以後、「UーA」と表記する）である。

[0029] 図29は、図28に示す匿名化前データセットが匿名化された、匿名化後データセットの一例を示す図である。図29に示す匿名化後データセットは、その匿名化前データセットの第1のレコードからなる第1のレコード群が、 $1=3$ で1ー多様性を充足するように匿名化されている。また、匿名化後データセットは、その匿名化前データセットの第2のレコードからなる第2のレコード群が、 $1=2$ で1ー多様性を充足するように匿名化されている。

[0030] 例えば、図28に示す匿名化前データセットにおいて、固有識別子が「6」、「7」及び「9」のレコードは、図29に示す匿名化後データセットにおいて、固有識別子に替えて同一のグループ識別子である「101」を付与されている。また、同一のグループ識別子を持つレコードは、準識別子である属性の属性値が同一の値に汎化されている。

[0031] 図28に示す匿名化前のデータセットにおいて、固有識別子が「6」、「7」及び「9」に対応する「同一の固有識別子を有するレコード間の対応関

係」は、「Y-E」、「X-D」及び「W-C」である。

[0032] 一方、図29に示す匿名化後データセットにおいて、グループ識別子が「101」に対応する「同一の固有識別子を有するレコード間の対応関係」は、「Y-E」、「Y-D」、「Y-C」、「X-E」、「X-D」、「X-C」、「W-E」、「W-D」及び「W-C」である。即ち、図29に示す匿名化後データセットは、図28に示す匿名化前のデータセットに存在しない「同一の固有識別子を有するレコード間の対応関係」である、「Y-C」及び「W-E」が余分に付加されてしまっている。

[0033] 以上が、「同一の固有識別子を有するレコード間の対応関係」の情報が曖昧になりすぎるという問題の具体例である。

[0034] 本発明の目的は、上述した問題点を解決する情報処理装置、匿名化方法、及びそのためのプログラムを提供することにある。

課題を解決するための手段

[0035] 本発明の情報処理装置は、固有識別子及び少なくとも1つの第1の属性を含む第1のレコードと、前記固有識別子と同一の固有識別子及び少なくとも1つの第2の属性を含む第2のレコードと、の組が複数件含まれるデータセットの中から、複数の前記第2のレコードを含む第2のレコード群において第2の1-多様性を充足可能であること、前記第2のレコード群に含まれる第2のレコードと組を成す前記第1のレコードから成る前記第1のレコード群において第1の1-多様性を充足可能であること、及び前記第1のレコードと前記第2のレコードとの間に存在する対応関係の抽象度に基づいて、複数の前記第2のレコードを抽出するレコード抽出手段と、前記レコード抽出手段によって抽出された前記第2のレコードからなる匿名グループデータセットを、前記匿名グループデータセットにおいて前記第2の1-多様性を充足可能であり、かつ前記匿名グループデータセットに含まれる第2のレコードと組を成す前記第1のレコードからなる第1のレコード群において前記第1の1-多様性を充足可能であるように、生成し、出力する匿名グループ生成手段と、を備える。

[0036] 本発明の匿名化方法は、コンピュータが、固有識別子及び少なくとも1つの第1の属性を含む第1のレコードと、前記固有識別子と同一の固有識別子及び少なくとも1つの第2の属性を含む第2のレコードと、の組が複数件含まれるデータセットの中から、前記第2のレコードからなる第2のレコード群において第2の1-多様性を充足可能であること、前記第2のレコード群に含まれる第2のレコードと組を成す前記第1のレコードから成る前記第1のレコード群において第1の1-多様性を充足可能であること、及び前記第1のレコードと前記第2のレコードとの間に存在する対応関係の抽象度に基づいて、複数の前記第2のレコードを抽出し、前記抽出された前記第2のレコードからなる匿名グループデータセットを、前記匿名グループデータセットにおいて前記第2の1-多様性を充足可能であり、かつ前記匿名グループデータセットに含まれる第2のレコードと組を成す前記第1のレコードからなる第1のレコード群において前記第1の1-多様性を充足可能であるように、生成し、出力する。

[0037] 本発明のコンピュータ読み取り可能な不揮発性記録媒体に記録されたプログラムは、固有識別子及び少なくとも1つの第1の属性を含む第1のレコードと、前記固有識別子と同一の固有識別子及び少なくとも1つの第2の属性を含む第2のレコードと、の組が複数件含まれるデータセットの中から、前記第2のレコードからなる第2のレコード群において第2の1-多様性を充足可能であること、前記第2のレコード群に含まれる第2のレコードと組を成す前記第1のレコードから成る前記第1のレコード群において第1の1-多様性を充足可能であること、及び前記第1のレコードと前記第2のレコードとの間に存在する対応関係の抽象度に基づいて、複数の前記第2のレコードを抽出する処理と、前記抽出された前記第2のレコードからなる匿名グループデータセットを、前記匿名グループデータセットにおいて前記第2の1-多様性を充足可能であり、かつ前記匿名グループデータセットに含まれる第2のレコードと組を成す前記第1のレコードからなる第1のレコード群において前記第1の1-多様性を充足可能であるように、生成し、出力する処

理と、をコンピュータに実行させる。

発明の効果

[0038] 本発明は、「同一の固有識別子（ユーザ識別子）を有するレコード間の対応関係」の情報を含むデータセットが、多様性を充足するように匿名化を施された場合に、その対応関係の情報が曖昧になりすぎることを防止することが可能になるという効果がある。

図面の簡単な説明

[0039] [図1]図1は、第1の実施形態に係る匿名化装置の構成を示すブロック図である。

[図2]図2は、第1の実施形態に係る匿名化装置を含むシステムの構成を示すブロック図である。

[図3]図3は、データセットの一例を示す図である。

[図4]図4は、ソートされた前提レコード分の一例を示す図である。

[図5]図5は、ソートされた結論レコード分の一例を示す図である。

[図6]図6は、前提匿名グループデータセットの一例を示す図である。

[図7]図7は、結論匿名グループデータセットの一例を示す図である。

[図8]図8は、抽出レコード群の一例を示す図である。

[図9]図9は、結論レコードを纏めた抽出結論レコード群の一例を表す図である。

[図10]図10は、共通部分レコード群の一例を示す図である。

[図11]図11は、前提属性値が同一の前提レコード毎に結論レコードを纏めた共通部分結論レコード群の一例を表す図である。

[図12]図12は、結論ソートレコード群の一例を示す図である。

[図13]図13は、結論属性値が同一の結論レコードを纏めた結論ソート結論レコード群の一例を表す図である。

[図14]図14は、匿名グループ結論レコード群の一例を示す図である。

[図15]図15は、グループ識別子毎に結論レコードを纏めた、匿名グループ結論レコード群の一例を表す図である。

[図16]図 1 6 は、本実施形態に係る匿名化装置を実現するコンピュータのハードウェア構成を示す図である。

[図17]図 1 7 は、本実施形態の動作を示すフローチャートである。

[図18]図 1 8 は、残レコードの一例を示す図である。

[図19]図 1 9 は、結論匿名グループの一例を示す図である。

[図20]図 2 0 は、結論匿名グループの一例を示す図である。

[図21]図 2 1 は、結論匿名グループの一例を示す図である。

[図22]図 2 2 は、第 2 の実施形態に係る匿名化装置 2 0 0 の構成を示すブロック図である。

[図23]図 2 3 は、遷移ベクトルの組み合わせの一例を示す図である。

[図24]図 2 4 は、2 つの遷移ベクトルの組み合わせの一例を示す図である。

[図25]図 2 5 は、遷移ベクトル間の類似度が「0」か否かを示す図である。

[図26]図 2 6 は、利用済みの遷移ベクトルを除いた遷移ベクトルの一例を示す図である。

[図27]図 2 7 は、遷移ベクトル間の組み合わせを示す図である。

[図28]図 2 8 は、匿名化前データセットの一例を示す図である。

[図29]図 2 9 は、匿名化後データセットの一例を示す図である。

発明を実施するための形態

[0040] 本発明を実施するための形態について図面を参照して詳細に説明する。尚、各図面及び明細書記載の各実施形態において、同様の構成要素には同様の符号を付与し、適宜説明を省略する。

[0041] <<<第 1 の実施形態>>>

図 1 は、本発明の第 1 の実施形態に係る匿名化装置 1 0 0 の構成を示すブロック図である。尚、匿名化装置（匿名化装置 1 0 0）は、一般的に情報処理装置とも呼ばれる。

[0042] 図 1 に示すように、本実施形態に係る匿名化装置 1 0 0 は、レコード抽出部 1 1 0 と、匿名グループ生成部 1 2 0 とを含む。

[0043] 図 2 は、本実施形態に係る匿名化装置 1 0 0 を含む匿名化システム 1 0 1

の構成を示すブロック図である。

[0044] 図2に示すように、匿名化システム101は、匿名化装置100と履歴情報記憶部500と匿名化情報記憶部600とを含む。

[0045] まず、匿名化システム101における、匿名化装置100の動作の概要を説明する。

[0046] ===履歴情報記憶部500===

履歴情報記憶部500は、図3に示すような、データセット510を記憶する。図3に示すように、例えば、データセット510は、固有識別子と、診療月、年齢及び病名の属性とを含む、複数のレコードから成る履歴情報である。また、データセット510は、同一の固有識別子を有する、「診療月」の属性値が「4月」のレコード（前提レコード）と「診療月」の属性値が「5月」のレコード（結論レコード）との間の対応関係の情報を含んでいる。

[0047] 前提レコードと結論レコードとは、同じ属性を含んでいなくてもよい。例えば、その前提レコードが固有識別子及びあるセンシティブ属性のみを含み、その結論レコードが固有識別子と他のセンシティブ属性のみを含む、データセットであってもよい。

[0048] 図4及び図5は、以下の説明の便宜上、図3に示すデータセット510を前提レコード（第1のレコード）分と結論レコード（第2のレコード）分に分けて示す図である。即ち、図4及び図5に示す前提レコード分521及び結論レコード分522は、匿名化装置100が生成するものではなく、説明のために便宜的に示す図である。図4は、前提レコードから成る、前提レコード分521を示す。図5は、結論レコードから成る、結論レコード分522を示す。

[0049] 以下の実施形態では、結論レコード分522を、前提レコード分521を参照しながら、前提レコード分521との間に存在する対応関係を保つように匿名化する方法について説明する。

[0050] ===匿名化装置100===

匿名化装置１００は、データセット５１０の中から、複数の結論レコード（結論レコード群、第１のレコード群とも呼ばれる）を抽出し、更に対応関係の抽象度に基づいて、その結論レコード群から複数の結論レコードを抽出する。ここで、その結論レコード群を構成するその複数の結論レコードは、その結論レコード群において第２の１－多様性を充足可能な複数の結論レコードであり、かつ、それらの結論レコードのそれぞれと組を成す複数の前提レコード（前提レコード群、第１のレコード群とも呼ばれる）において第１の１－多様性を充足可能であるような、複数の結論レコードである。

[0051] 次に、匿名化装置１００は、それらの抽出された複数の結論レコードから、結論レコードからなる結論匿名グループデータセット（匿名グループデータセットとも呼ばれる）を生成し、出力する。ここで、その結論レコードは、第２の１－多様性を充足し、かつ『それらの抽出された複数の結論レコード』との対応関係がある第１のレコード群に対して第１の１－多様性を充足する匿名化が可能なレコードである。

[0052] また、匿名化装置１００は、前提匿名グループデータセットに含まれる前提レコードのそれぞれ及び匿名グループデータセットに含まれる結論レコードのそれぞれに対して、それらの前提レコードと結論レコードとの対応関係を付与するようにしてもよい。ここで、その前提匿名グループデータセットは、結論匿名グループデータセットに含まれる結論レコードのそれぞれと組を成す複数の前提レコードが匿名化された、データセットである。

[0053] ===匿名化情報記憶部６００===

匿名化情報記憶部６００は、匿名化装置１００が出力する、前提匿名グループデータセット及び結論匿名グループデータセットを含む、匿名グループデータセットを記憶する。

[0054] 図６は、前提匿名グループデータセット６１１の一例を示す図である。図７は、結論匿名グループデータセット６１２の一例を示す図である。

[0055] 図６及び図７に示すように、前提匿名グループデータセット６１１及び結論匿名グループデータセット６１２のレコードのそれぞれは、固有識別子に

替えて、グループ識別子及び関連識別子を含む。尚、図6において、点線枠で囲ったその固有識別子は、前提レコード分521のレコードのそれぞれと、前提匿名グループデータセット611のレコードのそれぞれとの関連を判りやすくするために記載したものである。従って、その固有識別子は、前提匿名グループデータセット611には含まれない。尚、図7において点線枠で囲ったその固有識別子も、同様に結論匿名グループデータセット612には含まれない。

[0056] そのグループ識別子は、ある前提匿名グループに含まれる複数の前提レコードに同一に付与された識別子である。同様に、そのグループ識別子は、ある結論匿名グループに含まれる複数の結論レコードに同一に付与された識別子である。関連識別子は、同一の固有識別子を有する他方のレコードの、グループ識別子である。即ち、同一のそのグループ識別子に対応する複数の前提レコードは、一つの前提匿名グループを形成する。同様に、同一のグループ識別子に対応する複数の結論レコードは、一つの結論匿名グループを形成する。

[0057] 尚、前提匿名グループデータセット611及び結論匿名グループデータセット612のレコードのそれぞれは、これらの固有識別子を含んでもよい。その場合、匿名化情報記憶部600は、外部からの前提匿名グループデータセット611及び結論匿名グループデータセット612の取得要求に対して、それらの固有識別子を削除して出力するようにしてもよい。

[0058] 以上が、匿名化装置100の動作の概要の説明である。

[0059] 次に、匿名化装置100が備える各構成要素について、詳細に説明する。尚、図1に示す構成要素は、ハードウェア単位の構成要素でも、コンピュータ装置の機能単位に分割した構成要素でもよい。ここでは、図1に示す構成要素は、コンピュータ装置の機能単位に分割した構成要素として説明する。

[0060] ===レコード抽出部110===

レコード抽出部110は、遷移ベクトルを生成する。例えば、その遷移ベクトルは、前提レコードに含まれる第1の属性（以下、前提属性と呼ぶ）の

属性値毎の、結論レコードに含まれる第2の属性（以下、結論属性と呼ぶ）の各属性値が、その前提レコードと組を成す結論レコードに出現する頻度を要素とするベクトルである。換言すると、その遷移ベクトルは、前提属性の属性値毎の、結論属性の各属性値の出現頻度を要素とするベクトルである。ここで、前提属性は、前提レコードに含まれる第1の属性である。また、結論属性は、結論レコードに含まれる第2の属性である。その出現頻度は、結論属性の各属性値が、その前提レコードと組みを成す、その結論レコードに出現する頻度前提レコードと組みをなす。

[0061] 具体的には、レコード抽出部110は、図4に示す前提レコード分521及び図5に示す結論レコード分522を参照して次のように遷移ベクトルを算出する。

[0062] 前提レコードに含まれる前提属性は、図4に示す前提レコード分521の前提レコードの、病名の属性である。また、結論レコードに含まれる結論属性は、図5に示す結論レコード分522のレコードの、病名の属性である。

[0063] 例えば、病名の属性値が「U」のその前提レコードは、固有識別子が「1」、「13」、「27」、「39」、「14」、「26」、「28」、「29」、「38」、「11」及び「12」の前提レコード群のレコードである。これらの前提レコードと組を成す結論レコードは、同一の固有識別子「1」、「13」、「27」、「39」、「14」、「26」、「28」、「29」、「38」、「11」及び「12」を有する結論レコードである。

[0064] 次に、レコード抽出部110は、これらの結論レコードに含まれる病名の属性として出現する属性値の出現頻度を算出する。ここでその属性値は、4回出現の「A」、3回出現の「B」、2回出現の「C」及び2回出現の「D」である。従って、それぞれのその出現頻度は、「A」が0.37（ $=4 \div 11$ ）、「B」が0.28（ $=3 \div 11$ ）、「C」が0.19（ $=2 \div 11$ ）及び「D」が0.19（ $=2 \div 11$ ）である。また、その結論レコードの病名の属性の属性値である「E」及び「F」は、病名の属性値が「U」の前提レコードと組を成す結論レコードには出現しない。従って、「E」及び「

F」のそれぞれの出現頻度は、いずれも「0」である。

[0065] 以上より、レコード抽出部110は、その属性値の「U」についての遷移ベクトル t_{r_U} を、以下のとおり生成する。

[0066] $t_{r_U} = (0.37, 0.28, 0.19, 0.19, 0.00, 0.00)$
 $)^T$

同様にして、レコード抽出部110は、属性値の「V」、「W」、「X」、「Y」及び「Z」のそれぞれの、遷移ベクトル t_{r_V} 、 t_{r_W} 、 t_{r_X} 、 t_{r_Y} 及び t_{r_Z} を以下のとおり生成する。

[0067] $t_{r_V} = (0.22, 0.44, 0.22, 0.11, 0.00, 0.00)$
 $)^T$

$t_{r_W} = (0.22, 0.33, 0.33, 0.11, 0.00, 0.00)$
 $)^T$

$t_{r_X} = (0.20, 0.20, 0.00, 0.20, 0.40, 0.00)$
 $)^T$

$t_{r_Y} = (0.00, 0.00, 0.00, 0.67, 0.33, 0.00)$
 $)^T$

$t_{r_Z} = (0.00, 0.00, 0.00, 0.00, 0.00, 1.00)$
 $)^T$

次に、レコード抽出部110は、これらの遷移ベクトル間の類似度を算出する。レコード抽出部110は、それらの遷移ベクトルのいずれか2つの遷移ベクトルが、結論レコード群において第2の1ー多様性を充足可能である場合、それらの遷移ベクトル同士の類似度として、それらの遷移ベクトルの内積を算出する。尚、レコード抽出部110は、ベクトル間の類似性を表現する類似度、ベクトル間の非類似性を表現する距離であれば、内積に限らず、例えばユークリッド距離などを距離として算出してもよい。また、レコード抽出部110は、それらの遷移ベクトルのいずれか2つの遷移ベクトルが、結論レコード群において第2の1ー多様性を充足可能でない場合、それらの遷移ベクトル同士の類似度を「0」とする。

[0068] ここで、「2つの遷移ベクトルが、結論レコード群において第2の1ー多様性を充足可能」とは、2つの遷移ベクトルそれぞれに対応する結論レコードの、結論属性の結論属性値が、第2の1ー多様性の1種類（例えば、2種類）以上、共起していることである。即ちそれは、2つの遷移ベクトルそれぞれに対応する結論レコードのそれぞれ同士で同一である、結論属性の結論属性値が、第2の1ー多様性の1種類（例えば、2種類）以上、存在することである。

[0069] 具体的には、レコード抽出部110は、遷移ベクトル t_{r_U} と遷移ベクトル t_{r_V} との類似度 $s_{im}(U, V)$ を、遷移ベクトル t_{r_U} と遷移ベクトル t_{r_V} の内積である「0.26」と算出する。同様に、レコード抽出部110は、他の類似度を以下のとおり算出する。

[0070] $s_{im}(U, W) = 0.25$
 $s_{im}(U, X) = 0.16$
 $s_{im}(U, Y) = 0.12$
 $s_{im}(U, Z) = 0.00$
 $s_{im}(V, W) = 0.28$
 $s_{im}(V, X) = 0.16$
 $s_{im}(V, Y) = 0.07$
 $s_{im}(V, Z) = 0.00$
 $s_{im}(W, X) = 0.13$
 $s_{im}(W, Y) = 0.07$
 $s_{im}(W, Z) = 0.00$
 $s_{im}(X, Y) = 0.27$
 $s_{im}(X, Z) = 0.00$
 $s_{im}(Y, Z) = 0.00$

次に、レコード抽出部110は、類似度の大きい遷移ベクトルの順（即ち抽象度が小さい順）に、第1の1ー多様性の種類数のその遷移ベクトルに対応する、前提属性値を含む前提レコードと、その前提レコードと組を成す結

論レコードと、を抽出する。尚、「第1の1ー多様性の種類数のその遷移ベクトルに対応する」は、「対応関係を持つ前提レコード群（第2のレコードと組を成す第1のレコードから成る、第1のレコード群）において第1の1ー多様性を充足可能である」と言われることもある。

[0071] また、レコード抽出部110は、上述の結論レコードのみを抽出するようにしてもよい。この場合、レコード抽出部110は、以後の処理において、抽出した結論レコードの固有識別子に基づいて、データセット510の前提レコードを参照するようにしてもよい。

[0072] 具体的には、レコード抽出部110は、以下のようにして前提レコードと結論レコードとの組を抽出する。その抽出される前提レコードと結論レコードとの組は、抽象度が小さくなるように抽出されればよく、その順序はいかなる順序であってもよい。

[0073] ここで、前提レコードと結論レコードとの組の抽出の一例を示す。「U」、「V」、「W」、「X」及び「Y」のそれぞれの前提属性値に対応する、その類似度の合計のそれぞれは、「0.80」、「0.78」、「0.74」、「0.72」及び「0.54」である。そこで、レコード抽出部110は、その類似度の合計が最大である、「U」の前提属性値に対応する遷移ベクトル t_{r_u} 、を選択する。次に、レコード抽出部110は、遷移ベクトル t_{r_u} とのその類似度が大きい順に、遷移ベクトル t_{r_v} 及び遷移ベクトル t_{r_w} を選択する。

[0074] これらに対応する前提レコードと、その前提レコードと組を成す結論レコードとは、固有識別子が「1」、「13」、「27」、「39」、「14」、「26」、「28」、「29」、「38」、「11」、「12」、「2」、「25」、「10」、「15」、「16」、「30」、「24」、「31」、「3」、「32」、「37」、「4」、「22」、「23」、「9」、「17」、「36」及び「33」のレコードである。レコード抽出部110は、これらのレコードを抽出する。

[0075] 図8は、上述のようにしてレコード抽出部110が抽出した、抽出レコー

ド群 530 の一例を示す図である。図 8 は、組を成す前提レコード及び結論レコードのそれぞれが抽出前提レコード群 531 及び抽出結論レコード群 532 のそれぞれに含まれるレコードとして、抽出レコード群 530 を示す図である。

[0076] 図 9 は、図 8 に示す抽出レコード群 530 について、前提属性値が同一の前提レコード毎に、結論レコードを纏めた抽出結論レコード群 532 の一例を表す図である。尚、図 9 において、結論レコード 5321 の上段に固有識別子（例えば、「1」）を記し、下段に前提属性値と結論属性値と（例えば、「U-A」）を記す。以降の図 11、図 13、図 15、図 18、図 19、図 20 及び図 21 も同様である。

[0077] 図 9 に示すように、例えば、前提属性値が「U」の前提レコードに対応する、結論レコードは、固有識別子が「1」、「13」、「27」、「39」、「14」、「26」、「28」、「29」、「38」、「11」、「12」のレコードである。

[0078] ===匿名グループ生成部 120===

匿名グループ生成部 120 は、抽出レコード群 530 から、前提属性値が同一の前提レコード毎に、前提レコードと結論レコードとの組を抽出する。その抽出の際、匿名グループ生成部 120 は、その前提属性値が同一の前提レコードに対応する、結論属性値が同一の結論レコードの数が共通になるように、その前提レコードとその結論レコードとの組を抽出する。即ち、匿名グループ生成部 120 は、その前提属性値が同一の前提レコード毎に対応する、その結論属性値が同一のその結論レコードの数の最小値の分だけ、前提レコードと結論レコードとの組を抽出する。

[0079] また、匿名グループ生成部 120 は、上述の結論レコードのみを抽出するようにしてもよい。この場合、匿名グループ生成部 120 は、以後の処理において、抽出した結論レコードの固有識別子に基づいて、データセット 510 の前提レコードを参照するようにしてもよい。

[0080] 例えば、匿名グループ生成部 120 は、前提属性値が「U」、「V」及び

「W」それぞれの結論属性値が「A」の結論レコードの数を比較して、最小値が2であると判定する。

[0081] その最小値が2であることに基づいて、匿名グループ生成部120は、前提属性値が同一の前提レコード毎に、前提レコードと結論レコードとの組を2個ずつ抽出する。例えば、前提属性値が「U」である前提レコードと、それらに対する結論属性値が「A」である結論レコードとの組は、固有識別子が「1」、「13」、「27」及び「39」の、前提レコードと結論レコードとの組である。そこで、匿名グループ生成部120は、例えば、固有識別子が「1」及び「13」の、前提レコードと結論レコードとの組を抽出する。

[0082] 図10は、組を成す前提レコード及び結論レコードのそれぞれを共通部分前提レコード群541及び共通部分結論レコード群542のそれぞれに含まれるレコードとして、共通部分レコード群540の一例を示す図である。共通部分レコード群540は、図8に示す抽出レコード群530から抽出された、前提レコードと結論レコードとの組から成る。ここで、その前提レコードと結論レコードとは、前提属性値が同一の前提レコード毎に対応する結論レコード群が共通になるように抽出される。即ち、共通部分レコード群540は、前述のように抽出された前提レコード及び結論レコードのそれぞれを、共通部分前提レコード群541及び共通部分結論レコード群542として含む。

[0083] 図11は、図10に示す共通部分レコード群540について、前提属性値が同一の前提レコード毎に、結論レコードを纏めた共通部分結論レコード群542の一例を表す図である。

[0084] 図11に示すように、例えば、前提属性値が「U」、「V」及び「W」のそれぞれの前提レコードに対応する、結論属性値が「A」の結論レコードの数は、いずれも2つである。

[0085] 図12は、図10の共通部分レコード群540を、共通部分前提レコード群541の結論属性でソートした状態で、結論ソートレコード群550とし

て示す図である。図 1 2 に示す結論ソートレコード群 5 5 0 は、匿名化装置 1 0 0 が生成するものではなく、説明の便宜上示す図である。図 1 2 は、結論属性でソートされた状態の、組を成す前提レコードと結論レコードとの組のそれぞれが結論ソート前提レコード群 5 5 1 及び結論ソート結論レコード群 5 5 2 のそれぞれに含まれるレコードとして、結論ソートレコード群 5 5 0（共通部分レコード群 5 4 0）を示す。

[0086] 図 1 3 は、図 1 0 に示す共通部分結論レコード群 5 4 2 を、図 1 2 に示す結論ソート結論レコード群 5 5 2 にソートしたように、結論属性値が同一の結論レコードを纏めた結論ソート結論レコード群 5 5 2（共通部分結論レコード群 5 4 2）の一例を表す図である。

[0087] 図 1 3 に示すように、例えば、結論属性値が「A」の結論レコードは、前提属性値が「U」、「V」及び「W」のそれぞれの前提レコードに対応する 2 組の組み合わせ（以後、組み合わせ C と呼ぶ）を形成する。それらの 2 組の組み合わせ C は、固有識別子が、例えば「1」、「2」及び「3 2」の組み合わせと、「1 3」、「2 5」及び「3 7」の組み合わせとである。尚、組み合わせ C は、前提属性値が「U」、「V」及び「W」のそれぞれの前提レコードに対応する組み合わせであれば、任意の組み合わせであってよい。即ち、組み合わせ C は、第 1 の 1 ー多様性を充足する前提レコードに対応する組み合わせである。

[0088] 次に、匿名グループ生成部 1 2 0 は、共通部分結論レコード群 5 4 2 を利用して、第 2 の 1 ー多様性を満たす結論匿名グループにグループ分けされた結論レコードから成る、匿名グループ結論レコード群 5 6 2 を生成する。

[0089] 例えば、匿名グループ生成部 1 2 0 は、結論ソート結論レコード群 5 5 2 から、結論属性値が「B」の組み合わせ C と結論属性値が「A」の組み合わせ C とを選択して結論匿名グループを生成し、これにグループ識別子（例えば、「2 0 1」）を付与する。この際、匿名グループ生成部 1 2 0 は、組み合わせ C の残数が結論属性値毎にできるだけ均一になるように、組み合わせ C を選択してもよい。

[0090] 図14は、共通部分結論レコード群542を利用して生成された、匿名グループ結論レコード群562の一例を示す図である。尚、図中の点線枠で囲った前提レコード群は、結論レコードと前提レコードとの関連を判りやすくするために記載したものであり、匿名グループ結論レコード群562には含まれない。

[0091] 図15は、図14に示す匿名グループ結論レコード群562について、グループ識別子毎に結論レコードを纏めた、匿名グループ結論レコード群562の一例を表す図である。

[0092] 次に、匿名グループ生成部120は、匿名グループ結論レコード群562の各グループ（グループ識別子が同一の結論レコードの集合）毎に、結論属性以外の準識別子の属性値（ここでは、年齢の属性値）を汎化（同一の値に変換）し、図7に示す結論匿名グループデータセット612を生成し、結論匿名グループデータセット（第2の匿名グループデータセット）として出力する。尚、図7に示す結論匿名グループデータセット612はグループ識別子でソートされているが、匿名グループ生成部120が出力する結論匿名グループデータセットの結論レコードは任意の並び順であってよい。

[0093] 尚、匿名グループ生成部120は、結論属性以外の準識別子の属性値（ここでは、診療月及び年齢の属性値）を汎化する必要がない場合（例えば、結論レコードがこれらの属性を含んでいない場合）、匿名グループ結論レコード群562をそのまま、結論匿名グループデータセットとして出力するようにしてもよい。

[0094] 以上は、結論レコードからなる結論匿名グループデータセットの生成についての説明である。

[0095] 次に、前提レコードからなる前提匿名グループデータセットの生成について説明する。尚、その前提匿名グループデータセットは、以下の方法に限らず、他の匿名化装置や方法によって生成されていてもよい。

[0096] 匿名グループ生成部120は、図10に示す共通部分前提レコード群541を利用して、図6に示す前提匿名グループデータセット611を生成し、

出力する。

[0097] 具体的には、匿名グループ生成部120は、共通部分前提レコード群541の先頭から第1の1ー多様性の種類数の前提属性値に対応する前提レコードの組み合わせ（例えば、固有識別子が「1」、「2」及び「32」の前提レコードの組み合わせ）を順次抽出する。そして、匿名グループ生成部120は、その抽出した組み合わせのそれぞれに、グループ識別子（例えば、「101」）を付与する。即ち、その抽出した組み合わせのそれぞれは、前提匿名グループを形成する。

[0098] 次に、匿名グループ生成部120は、同一のグループ識別子を付与した前提レコードのそれぞれの、前提属性以外の準識別子の属性値（ここでは、年齢の属性値）を汎化（同一の値に変換）する。

[0099] 更に、匿名グループ生成部120は、同一の固有識別子を有する結論レコードのグループ識別子を関連識別子とし、図6に示す前提匿名グループデータセット611を生成する。

[0100] 以上が、前提レコードからなる前提匿名グループデータセットの生成についての説明である。

[0101] 以上が、匿名化装置100の機能単位の各構成要素についての説明である。

[0102] 次に、匿名化装置100のハードウェア単位の構成要素について説明する。

[0103] 図16は、本実施形態に係る匿名化装置100を実現するコンピュータ700のハードウェア構成を示す図である。

[0104] 図16に示すように、コンピュータ700は、CPU（Central Processing Unit）701、記憶部702、記憶装置703、入力部704、出力部705及び通信部706を含む。更に、コンピュータ700は、外部から供給される記録媒体（または記憶媒体）707を含む。記録媒体707は、情報を非一時的に記憶する不揮発性記録媒体であってもよい。

- [0105] CPU 701は、オペレーティングシステム（不図示）を動作させて、コンピュータ 700の、全体の動作を制御する。また、CPU 701は、例えば記憶装置 703に装着された記録媒体 707から、プログラムやデータを読み込み、読み込んだプログラムやデータを記憶部 702に書き込む。ここで、そのプログラムは、例えば、後述の図 17に示すフローチャートの動作をコンピュータ 700に実行させるプログラムである。
- [0106] そして、CPU 701は、読み込んだプログラムに従って、また読み込んだデータに基づいて、図 1に示すレコード抽出部 110及び匿名グループ生成部 120として各種の処理を実行する。
- [0107] 尚、CPU 701は、通信網（不図示）に接続されている外部コンピュータ（不図示）から、記憶部 702にプログラムやデータをダウンロードするようにしてもよい。
- [0108] 記憶部 702は、プログラムやデータを記憶する。記憶部 702は、データセット 510、抽出レコード群 530、共通部分レコード群 540、匿名グループ結論レコード群 562、前提匿名グループデータセット 611及び結論匿名グループデータセット 612などを記憶してもよい。また、記憶部 702は、履歴情報記憶部 500及び匿名化情報記憶部 600を含んでよい。
- [0109] 記憶装置 703は、例えば、光ディスク、フレキシブルディスク、磁気光ディスク、外付けハードディスク及び半導体メモリであって、記録媒体 707を含む。記憶装置 703（記録媒体 707）は、プログラムをコンピュータ読み取り可能に記憶する。また、記憶装置 703は、データを記憶してもよい。記憶装置 703は、記憶部 702と同様のデータを記憶してもよい。また、記憶装置 703は、履歴情報記憶部 500及び匿名化情報記憶部 600を含んでよい。
- [0110] 入力部 704は、例えばマウスやキーボード、内蔵のキーボタンなどで実現され、入力操作に用いられる。入力部 704は、マウスやキーボード、内蔵のキーボタンに限らず、例えばタッチパネル、加速度計、ジャイロセンサ

、カメラなどでもよい。

[0111] 出力部 705 は、例えばディスプレイで実現され、出力を確認するために用いられる。

[0112] 通信部 706 は、外部とのインタフェースを実現する。通信部 706 は、レコード抽出部 110 及び匿名グループ生成部 120 の一部として含まれる。

[0113] 以上説明したように、図 1 に示す匿名化装置 100 の機能単位のブロックは、図 2 に示すハードウェア構成のコンピュータ 700 によって実現される。但し、コンピュータ 700 が備える各部の実現手段は、上記に限定されない。すなわち、コンピュータ 700 は、物理的に結合した 1 つの装置により実現されてもよいし、物理的に分離した 2 つ以上の装置を有線または無線で接続し、これら複数の装置により実現されてもよい。

[0114] 尚、上述のプログラムのコードを記録した記録媒体 707 が、コンピュータ 700 に供給され、CPU 701 は、記録媒体 707 に格納されたプログラムのコードを読み出して実行するようにしてもよい。或いは、CPU 701 は、記録媒体 707 に格納されたプログラムのコードを、記憶部 702、記憶装置 703 またはその両方に格納するようにしてもよい。すなわち、本実施形態は、コンピュータ 700 (CPU 701) が実行するプログラム (ソフトウェア) を、一時的にまたは非一時的に、記憶する記録媒体 707 の実施形態を含む。

[0115] 以上が、本実施形態における匿名化装置 100 を実現するコンピュータ 700 の、ハードウェア単位の各構成要素についての説明である。

[0116] 次に本実施形態の動作について、図 1 ~ 図 17 を参照して詳細に説明する。

[0117] 図 17 は、本実施形態の動作を示すフローチャートである。尚、このフローチャートによる処理は、前述した CPU によるプログラム制御に基づいて、実行されても良い。また、処理のステップ名については、S601 のように、記号で記載する。

- [0118] レコード抽出部 110 は、遷移ベクトルを生成する（S601）。
- [0119] 次に、レコード抽出部 110 は、遷移ベクトル間の類似度を算出する（S602）。
- [0120] 次に、レコード抽出部 110 は、類似度の大きい遷移ベクトルの順に、第 1 の 1ー多様性の種類数のその遷移ベクトルに対応する、前提属性値を含む前提レコードと、その前提レコードと組を成す結論レコードと、を抽出し、抽出レコード群 530 として出力する（S603）。
- [0121] 次に、匿名グループ生成部 120 は、抽出レコード群 530 から、前提属性値が同一の前提レコード毎に、「それらの前提レコードに対応する、結論属性値が同一の結論レコードの数が共通」になるように、共通部分レコード群 540 として、前提レコードと結論レコードとの組を抽出する（S604）。
- [0122] 次に、匿名グループ生成部 120 は、共通部分結論レコード群 542 を利用して、第 2 の 1ー多様性を満たす結論匿名グループにグループ分けされた結論レコードから成る匿名グループ結論レコード群 562 を生成する（S606）。
- [0123] 次に、匿名グループ生成部 120 は、匿名グループ結論レコード群 562 のグループ毎に、結論属性以外の準識別子の属性値を汎化し、結論匿名グループデータセット 612 を生成し、結論匿名グループとして出力する（S607）。
- [0124] 次に、匿名グループ生成部 120 は、前提レコードのグループ化を行う。匿名グループ生成部 120 は、共通部分前提レコード群 541 の先頭から第 1 の 1ー多様性の種類数の前提属性値に対応する前提レコードの組み合わせを順次抽出し、その抽出した組み合わせのそれぞれにグループ識別子を付与する（S608）。
- [0125] 但し、前提レコードのグループ化は、この方法によらず、様々な方法を用いてよい。例えば、ここでの前提レコードを結論レコードとし、他のレコード群を前提レコードとして、ここでの前提レコードがグループ化されていて

もよい。

[0126] 次に、匿名グループ生成部 120 は、同一のグループ識別子を付与した前提レコードそれぞれの、前提属性以外の準識別子の属性値を汎化する（S609）。

[0127] 次に、匿名グループ生成部 120 は、同一の固有識別子を有する結論レコードのグループ識別子を関連識別子として、図 6 に示す前提匿名グループデータセット 611 を生成し、出力する（S610）。

[0128] <<<本実施形態の第 1 の変形例>>>

匿名グループ生成部 120 は、図 17 に示す動作において出力された前提匿名グループデータセット（第 1 の匿名グループデータセット）と結論匿名グループデータセット（第 2 の匿名グループデータセット）とに、対応関係の抽象化が発生しないように追加可能な、残レコードを追加する。ここで、その残レコードは、その結論匿名グループデータセットに含まれる結論レコードの有する固有識別子以外の、他の固有識別子を有する結論レコードである。

[0129] 図を用いて、具体的な例を説明する。

[0130] 図 18 は、図 5 に示す結論レコード分 522 から、図 7 に示す結論匿名グループデータセット 612 を除いた残レコード 570 の一例を示す図である。

[0131] 匿名グループ生成部 120 は、特定の結論匿名グループに対して、以下の条件に合致する、複数の前提レコードと結論レコードとの組を追加する。第 1 の条件は、その複数の前提レコードが、その特定の結論匿名グループに含まれる結論レコードと組を成す、いずれの前提レコードの前提属性値とも異なる、同一の前提属性値を有することである。第 2 の条件は、その複数の結論レコードが、その特定の結論匿名グループに含まれる前提レコードのそれぞれの、前提属性値の全ての種類を含むことである。

[0132] 例えば、匿名グループ生成部 120 は、図 17 に示すステップ S606 の次に、特定の結論匿名グループとして、グループ識別子が「201」のグル

ープを選択する。

[0133] 更に、匿名グループ生成部120は、前提属性値が「U」、「V」及び「W」以外の前提属性値に対応し、結論属性値が「A」及び「B」を有する結論レコードを残レコード570から抽出する。

[0134] 次に、匿名グループ生成部120は、抽出した結論レコードに「201」のグループ識別子を付与する。

[0135] 次に、匿名グループ生成部120は、図7に示すステップS607以降の処理を、抽出した結論レコードとそれに対応する前提レコードを含めて、実行する。

[0136] 図19は、上述のようにして形成された、グループ識別子が「201」の結論匿名グループの一例を、模式的に示す図である。図19に示すように、匿名化前の固有識別子毎の対応関係は8種類である。また、これらの結論レコードが全て同じグループ識別子のもとにグループ化された場合、即ち前提属性値と結論属性値とを任意に入れ替え可能とされた場合、やはり対応関係は8種類である。即ち、対応関係の抽象化は発生しない。

[0137] また、匿名グループ生成部120は、特定の結論匿名グループに対して、以下の条件に合致する、複数の前提レコードと結論レコードとの組を追加するようにしてもよい。第一の条件は、その複数の結論レコードが、その特定の結論匿名グループに含まれる結論レコードの結論属性値のいずれとも異なる、同一の結論属性値を有することである。第2の条件は、その複数の前提レコードのそれぞれが、その特定の結論匿名グループに含まれる結論レコードに対応する前提レコードのそれぞれの、前提属性値の全ての種類を含むことである。

[0138] 図20は、上述の条件に基づいて形成された結論匿名グループの一例を模式的に示す図である。

[0139] <<<本実施形態の第2の変形例>>>

匿名グループ生成部120は、残レコードから、第1の1ー多様性及び第2の1ー多様性のそれぞれを充足する匿名化が可能な、前提レコードからな

る前提匿名グループ及び結論レコードからなる結論匿名グループのそれぞれを生成する。ここで、その残レコードは、図 17 に示す動作において出力された結論匿名グループデータセットに含まれる結論レコードの有する固有識別子以外の、固有識別子を有する結論レコードである。

[0140] 図 21 は、残レコード 570 から生成した結論匿名グループの一例を示す図である。図 21 に示すように、上述のようにして生成された結論匿名グループは、第 2 の 1-多様性を充足し、それらの結論レコードに対応する前提レコードからなる匿名グループは、第 1 の 1-多様性を充足する。但し、匿名化前の固有識別子毎の対応関係は 5 種類であるのに対し、グループ化された場合の対応関係は 9 種類である。従って、対応関係の抽象化が、発生する。

[0141] <<<本実施形態の第 3 の変形例>>>

上述の説明においては、レコード抽出部 110 及び匿名グループ生成部 120 は、診療月の属性値が「4 月」のレコードを前提レコード（第 1 のレコード）とし、診療月の属性値が「5 月」のレコードを結論レコード（第 2 のレコード）として、処理した。しかし、レコード抽出部 110 及び匿名グループ生成部 120 は、診療月の属性値が「5 月」のレコードを前提レコード（第 1 のレコード）とし、診療月の属性値が「4 月」のレコードを結論レコード（第 2 のレコード）としてもよい。

[0142] 即ち、対応関係は、属性の物理的な性質に係わらず、任意の方向の対応関係であってよい。

[0143] <<<本実施形態の第 4 の変形例>>>

上述の説明においては、レコード抽出部 110 及び匿名グループ生成部 120 は、各動作におけるレコードの抽出及び選択を、前提属性値と結論属性値との関係のみを考慮して、図示されている順番で行うようにした。しかし、レコード抽出部 110 及び匿名グループ生成部 120 は、他の属性の匿名化（例えば、年齢の汎化）を考慮して、各動作におけるレコードの抽出及び選択を行う（例えば、年齢の属性値が近いレコードを同一のグループにする

) ようにしてもよい。

[0144] <<<本実施形態の第5の変形例>>>

図7に示すステップS608からステップS610までの処理のそれぞれは、その順番を守った上で、ステップS604以降の、任意のタイミングで実行してもよい。

[0145] <<<本実施形態の第6の変形例>

匿名グループ生成部120は、前提匿名グループデータセットと結論匿名データセットとを別々に出力してもよいし、纏めて1つのデータセットとして出力してもよい。

[0146] <<<本実施形態の第7の変形例>>>

匿名グループ生成部120は、結論匿名グループデータセットの結論レコードに対して、対応する前提レコードのグループ識別子を、関連識別子として関連付けてもよい。この場合、匿名グループ生成部120は、前提レコードに関連識別子を関連付けないようにしてもよい。

[0147] <<<本実施形態の第8の変形例>>>

匿名グループ生成部120は、対応関係にある前提匿名グループの前提レコードと結論匿名グループの結論レコードとについてグループ識別子を一致させてもよい。この場合、匿名グループ生成部120は、前提レコード及び結論レコードに関連識別子を関連付けないようにしてもよい。

[0148] 上述した本実施形態における第1の効果は、「同一の固有識別子を有するレコード間の対応関係」の情報を含むデータセットが1ー多様性を充足するように匿名化を施された場合に、その対応関係の情報が曖昧になりすぎることを防止することが可能になる点である。

[0149] その理由は、以下のような構成を含むからである。即ち、第1にレコード抽出部110が、第1及び第2の1ー多様性を充足可能であることと対応関係の抽象度とに基づいて、前提レコード及び結論レコードを抽出する。第2に、匿名グループ生成部120が、レコード抽出部110によって抽出された前提レコードを参照し、同じく抽出された結論レコードから第1の1ー多

様性と第2の1-多様性とを充足可能にするように結論レコードを抽出して結論匿名グループを生成する。

[0150] 上述した本実施形態における第2の効果は、「同一の固有識別子を有するレコード間の対応関係」の情報を含むデータセットが、前提レコードと結論レコードとで異なる1の値の1-多様性を充足するように匿名化を施された場合にも、その対応関係の情報が曖昧になりすぎることを防止することが可能になる点である。

[0151] その理由は、第1の効果の理由と同じである。

[0152] 上述した本実施形態における第3の効果は、データセットに含まれるレコードをより有効に利用することが可能になる点である。

[0153] その理由は、匿名グループ生成部120が、前提匿名グループデータセットと結論匿名グループデータセットとに、対応関係の抽象化が発生しないように追加可能な、残レコードを追加するようにしたからである。

[0154] 上述した本実施形態における第4の効果は、データセットに含まれるレコードを、更に、より有効に利用することが可能になる点である。

[0155] その理由は、匿名グループ生成部120が、残レコードから、前提匿名グループ及び結論匿名グループのそれぞれを生成するようにしたからである。

[0156] 上述した本実施形態における第5の効果は、データセットの匿名化を、利用価値が低くならないように施すことが可能になる点である。

[0157] その理由は、レコード抽出部110及び匿名グループ生成部120が、他の属性の匿名化を考慮して、各動作におけるレコードの抽出及び選択を、行うようにしたからである。

[0158] <<<第2の実施形態>>>

次に、本発明の第2の実施形態について図面を参照して詳細に説明する。以下、本実施形態の説明が不明確にならない範囲で、前述の説明と重複する内容については説明を省略する。

[0159] 図22は、本発明の第2の実施形態に係る匿名化装置200の構成を示すブロック図である。

[0160] 図 2 2 に示す構成要素は、ハードウェア単位の構成要素ではなく、機能単位の構成要素を示している。尚、図 2 2 に示す構成要素は、ハードウェア単位の構成要素でも、コンピュータ装置の機能単位に分割した構成要素でもよい。ここでは、図 1 に示す構成要素は、コンピュータ装置の機能単位に分割した構成要素として説明する。

[0161] 図 2 2 を参照すると、本実施形態に係る匿名化装置 2 0 0 は、第 1 の実施形態の匿名化装置 1 0 0 と比べて、遷移ベクトル抽出部 2 3 0 を更に含み、レコード抽出部 1 1 0 に替えてレコード抽出部 2 1 0 を含む。

[0162] === 遷移ベクトル抽出部 2 3 0 ===

遷移ベクトル抽出部 2 3 0 は、複数の遷移ベクトルについての類似度の算出対象を示す、算出対象情報を生成する。そして、遷移ベクトル抽出部 2 3 0 は、その算出対象情報をレコード抽出部 2 1 0 に出力する。

[0163] 算出対象情報に含まれる算出対象を抽出する操作について具体的に説明する。

[0164] <<<第一の抽出操作>>>

遷移ベクトル抽出部 2 3 0 は、2 つの遷移ベクトル間に、第 2 の I - 多様性の I 種類以上の、要素の共起が存在する場合、その 2 つの遷移ベクトルの組み合わせを算出対象として抽出する。

[0165] 例えば、第 2 の I - 多様性の I が「2」であるとする。また、遷移ベクトル抽出部 2 3 0 が処理の対処とする複数の遷移ベクトルを以下のとおりであるとする。

[0166]
$$\begin{aligned} t r_A &= (0.3, 0.2, 0.2, 0.0, 0.0, 0.0, \\ &0.0, 0.1, 0.1, 0.0, 0.2)^T \\ t r_B &= (0.2, 0.0, 0.2, 0.0, 0.0, 0.0, \\ &0.0, 0.0, 0.1, 0.3, 0.2)^T \\ t r_C &= (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, \\ &0.0, 0.1, 0.1, 0.2, 0.0)^T \\ t r_D &= (0.0, 0.0, 0.1, 0.0, 0.2, 0.1, \end{aligned}$$

$$0.1, 0.2, 0.2, 0.0, 0.0)^T$$

$$t_{r_E} = (0.0, 0.0, 0.2, 0.1, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)^T$$

$$t_{r_F} = (0.0, 0.0, 0.0, 0.0, 0.0, 0.1, 0.2, 0.0, 0.0, 0.0, 0.0)^T$$

$$t_{r_G} = (0.0, 0.0, 0.1, 0.2, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)^T$$

この場合、遷移ベクトル t_{r_A} と遷移ベクトル t_{r_B} とは、1、3、9 及び 11 番目の各要素が共起している。従って、遷移ベクトル抽出部 230 は、遷移ベクトル t_{r_A} と遷移ベクトル t_{r_B} との組み合わせを算出対象として抽出する。

[0167] また、遷移ベクトル t_{r_A} と遷移ベクトル t_{r_E} とは、3 番目の要素だけが共起している（共起している要素が 1 種類である）。従って、遷移ベクトル抽出部 230 は、遷移ベクトル t_{r_A} と遷移ベクトル t_{r_E} との組み合わせを算出対象として抽出しない。

[0168] 図 23 は、遷移ベクトル抽出部 230 が抽出した 2 つの遷移ベクトルの組み合わせの、一例を示す図である。図 23 は、各遷移ベクトルをノードとし、算出対象の 2 つのベクトルの組み合わせをエッジで示す。

[0169] 以上のようにして、遷移ベクトル抽出部 230 は、例えば、以下に示す算出対象情報を生成する。

[0170] $(t_{r_A} - t_{r_B}, t_{r_A} - t_{r_C}, t_{r_A} - t_{r_D}, t_{r_B} - t_{r_C}, t_{r_B} - t_{r_D}, t_{r_C} - t_{r_D}, t_{r_D} - t_{r_E}, t_{r_D} - t_{r_G}, t_{r_D} - t_{r_F}, t_{r_E} - t_{r_G})$

<<<第二の抽出操作>>>

遷移ベクトル抽出部 230 は、ある遷移ベクトルについて、その遷移ベクトルとの類似度が「0」ではない他の遷移ベクトルが、第 1 の I - 多様性の I 種類の「 $I - 1$ 」個以上存在する場合、その遷移ベクトルと他の遷移ベクトルとの組み合わせを算出対象として抽出する。

- [0171] 尚、遷移ベクトル抽出部230は、類似度が遷移ベクトル間の内積である場合、それらの遷移ベクトルに対応する各要素間のそれぞれの、論理積をとることで遷移ベクトル間の類似度が「0」か否かを判定する。即ち、各要素間の論理積の全てが「0」の場合、遷移ベクトル抽出部230は、遷移ベクトル間の類似度が「0」であると判定する。各要素間の論理積のいずれかが「0」でない場合、遷移ベクトル抽出部230は、遷移ベクトル間の類似度が「0」でないと判定する。
- [0172] 例えば、第1の1-多様性の1が「3」であるとする。また、遷移ベクトル抽出部230が処理の対処とする複数の遷移ベクトルを第一の抽出操作で示した例のとおりであるとする。
- [0173] この場合、遷移ベクトル t_{r_A} について、遷移ベクトル t_{r_A} との類似度が「0」でない他の遷移ベクトルは遷移ベクトル t_{r_B} 、遷移ベクトル t_{r_C} 及び遷移ベクトル t_{r_D} である。従って、遷移ベクトル抽出部230は、遷移ベクトル t_{r_A} と遷移ベクトル t_{r_B} 及び遷移ベクトル t_{r_C} との組み合わせを算出対象として抽出する。
- [0174] また、遷移ベクトル t_{r_F} について、遷移ベクトル t_{r_F} との類似度が「0」でない他の遷移ベクトルは遷移ベクトル t_{r_D} のみである。従って、遷移ベクトル抽出部230は、遷移ベクトル t_{r_F} と他の遷移ベクトルとの組み合わせを算出対象として抽出しない。
- [0175] 図24は、遷移ベクトル抽出部230が抽出した2つの遷移ベクトルの組み合わせの、一例を示す図である。図24は、各遷移ベクトルをノードとし、算出対象の2つのベクトルの組み合わせをエッジで示す。
- [0176] 以上のようにして、遷移ベクトル抽出部230は、例えば、以下に示す算出対象情報を生成する。
- [0177] ($t_{r_A} - t_{r_B}$ 、 $t_{r_A} - t_{r_C}$ 、 $t_{r_A} - t_{r_D}$ 、 $t_{r_B} - t_{r_C}$ 、 $t_{r_B} - t_{r_D}$ 、 $t_{r_C} - t_{r_D}$ 、 $t_{r_D} - t_{r_E}$ 、 $t_{r_D} - t_{r_G}$ 、 $t_{r_E} - t_{r_G}$)

<<<第三の抽出操作>>>

遷移ベクトル抽出部230は、第1の1-多様性の1個のある遷移ベクトルについて、それらの遷移ベクトル間の類似度のいずれもが、「0」ではない場合、それらの遷移ベクトル間の組み合わせを算出対象として抽出する。

[0178] 図25は、遷移ベクトル抽出部230が処理対象とする遷移ベクトル間の類似度が「0」か否かを示す模式図である。図25は、各遷移ベクトルをノードとし、ある2つの遷移ベクトル間の類似度が「0」でないことをエッジで示す。

[0179] 例えば、第1の1-多様性の1が「3」の場合、遷移ベクトル抽出部230は、3個の遷移ベクトル t_{r_A} 、遷移ベクトル t_{r_B} 及び遷移ベクトル t_{r_C} について、それらの遷移ベクトル間の類似度のいずれもが、「0」ではない（エッジがある）ので、それらの遷移ベクトル間の組み合わせを算出対象として抽出する。また、遷移ベクトル抽出部230は、3個の遷移ベクトル t_{r_D} 、遷移ベクトル t_{r_E} 及び遷移ベクトル t_{r_F} について、遷移ベクトル t_{r_D} と遷移ベクトル t_{r_F} との類似度が「0」であるので、それらの遷移ベクトル間の組み合わせを、算出対象として抽出しない。

[0180] 以上のようにして、遷移ベクトル抽出部230は、例えば、以下に示す算出対象情報を生成する。

[0181] ($t_{r_A}-t_{r_B}$ 、 $t_{r_A}-t_{r_C}$ 、 $t_{r_A}-t_{r_D}$ 、 $t_{r_B}-t_{r_C}$ 、 $t_{r_B}-t_{r_D}$ 、 $t_{r_C}-t_{r_D}$ 、 $t_{r_F}-t_{r_G}$ 、 $t_{r_F}-t_{r_H}$ 、 $t_{r_G}-t_{r_H}$)

また、同様にして、第1の1-多様性の1が「4」の場合、遷移ベクトル抽出部230は、以下に示す算出対象情報を生成する。

[0182] ($t_{r_A}-t_{r_B}$ 、 $t_{r_A}-t_{r_C}$ 、 $t_{r_A}-t_{r_D}$ 、 $t_{r_B}-t_{r_C}$ 、 $t_{r_B}-t_{r_D}$ 、 $t_{r_C}-t_{r_D}$)

以上が、算出対象情報に含まれる算出対象を抽出する操作の説明である。

[0183] 尚、遷移ベクトル抽出部230は、上述の第一、二及び三の抽出操作を単独でも、任意に組み合わせてでも、実行してよい。

[0184] ===レコード抽出部210===

レコード抽出部 210 は、生成した遷移ベクトルを遷移ベクトル抽出部 230 に出力する。そして、レコード抽出部 210 は、遷移ベクトル抽出部 230 からその抽出した結果を受け取る。

[0185] 例えば、レコード抽出部 210 は、図 17 に示すステップ S601 に続けて、生成した遷移ベクトルを遷移ベクトル抽出部 230 に出力する。そして、レコード抽出部 210 は、遷移ベクトル抽出部 230 からその抽出した結果を受け取ると、ステップ S602 以後の動作を実行する。

[0186] 尚、レコード抽出部 210 は、図 17 に示すステップ S603 に続けて、利用済みの遷移ベクトルを除いた遷移ベクトルを遷移ベクトル抽出部 230 に出力するようにしてもよい。この場合、レコード抽出部 210 は、遷移ベクトル抽出部 230 からその抽出した結果を受け取ると、再度ステップ S602 から以後の動作を実行するようにしてもよい。ここで、利用済みの遷移ベクトルは、ステップ S603 において抽出した前提レコードに対応する遷移ベクトルである。

[0187] 図 26 は、レコード抽出部 210 が出力する、利用済みの遷移ベクトルを除いた遷移ベクトルの一例を示す図である。例えば、レコード抽出部 210 は、図 17 のステップ S603 において、3 個の遷移ベクトル t_{r_A} 、遷移ベクトル t_{r_B} 及び遷移ベクトル t_{r_C} を利用したとする。この場合、レコード抽出部 210 は、3 個の遷移ベクトル t_{r_A} 、遷移ベクトル t_{r_B} 及び遷移ベクトル t_{r_C} を除いた、遷移ベクトル t_{r_D} 、遷移ベクトル t_{r_E} 、遷移ベクトル t_{r_G} 及び遷移ベクトル t_{r_H} を遷移ベクトル抽出部 230 に出力する。

[0188] 図 27 は、遷移ベクトル抽出部 230 が、レコード抽出部 210 から受け取った遷移ベクトルについて、算出対象として抽出する遷移ベクトル間の組み合わせを示す図である。この場合、遷移ベクトル抽出部 230 は、以下に示す算出対象情報を生成する。

[0189] ($t_{r_D} - t_{r_E}$ 、 $t_{r_D} - t_{r_G}$ 、 $t_{r_E} - t_{r_G}$)

上述した本実施形態における第 1 の効果は、第 1 の実施形態の効果に加えて、効率よく匿名化することが可能になる点である。

- [0190] その理由は、遷移ベクトル抽出部 230 が複数の遷移ベクトルについての類似度の算出対象を示す算出対象情報を生成し、レコード抽出部 210 がその算出対象情報に基づいて、類似度を算出するようにしたからである。即ち、必要のない類似度について、その算出処理を実行しないようにしたからである。
- [0191] また、レコード抽出部 210 が、利用済みの遷移ベクトルを除いた遷移ベクトルを遷移ベクトル抽出部 230 に出力し、算出対象情報を取得するようにしたので、更に匿名化を効率化することが可能になる。
- [0192] 以上の各実施形態で説明した各構成要素は、必ずしも個々に独立した存在である必要はない。例えば、各構成要素は、複数の構成要素が 1 個のモジュールとして実現されてよい。また、各構成要素は、1 つの構成要素が複数のモジュールで実現されてもよい。また、各構成要素は、ある構成要素が他の構成要素の一部であるような構成であってよい。また、各構成要素は、ある構成要素の一部と他の構成要素の一部とが重複するような構成であってもよい。
- [0193] 以上説明した各実施形態における各構成要素及び各構成要素を実現するモジュールは、必要に応じ、可能であれば、ハードウェア的に実現されてよい。また、各構成要素及び各構成要素を実現するモジュールは、コンピュータ及びプログラムで実現されてよい。また、各構成要素及び各構成要素を実現するモジュールは、ハードウェア的なモジュールとコンピュータ及びプログラムとの混在により実現されてもよい。
- [0194] そのプログラムは、例えば、磁気ディスクや半導体メモリなど、不揮発性のコンピュータ可読記録媒体に記録されて提供され、コンピュータの立ち上げ時などにコンピュータに読み取られる。この読み取られたプログラムは、そのコンピュータの動作を制御することにより、そのコンピュータを前述した各実施形態における構成要素として機能させる。
- [0195] また、以上説明した各実施形態では、複数の動作をフローチャートの形式で順番に記載してあるが、その記載の順番は複数の動作を実行する順番を限

定するものではない。このため、各実施形態を実施するときには、その複数の動作の順番は内容的に支障しない範囲で変更することができる。

[0196] 更に、以上説明した各実施形態では、複数の動作は個々に相違するタイミングで実行されることに限定されない。例えば、ある動作の実行中に他の動作が発生したり、ある動作と他の動作との実行タイミングが部分的に乃至全部において重複していたりしていてもよい。

[0197] 更に、以上説明した各実施形態では、ある動作が他の動作の契機になるように記載しているが、その記載はある動作と他の動作との全ての関係を限定するものではない。このため、各実施形態を実施するときには、その複数の動作の関係は内容的に支障のない範囲で変更することができる。また各構成要素の各動作の具体的な記載は、各構成要素の各動作を限定するものではない。このため、各構成要素の具体的な各動作は、各実施形態を実施する上で機能的、性能的、その他の特性に対して支障をきたさない範囲内で変更されて良い。

[0198] 以上、各実施形態を参照して本発明を説明したが、本発明は上記実施形態に限定されるものではない。本発明の構成や詳細には、本発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

[0199] この出願は、2012年9月26日に提出された日本出願特願2012-212454を基礎とする優先権を主張し、その開示の全てをここに取り込む。

符号の説明

- [0200]
- | | |
|-----|-----------|
| 100 | 匿名化装置 |
| 101 | 匿名化システム |
| 110 | レコード抽出部 |
| 120 | 匿名グループ生成部 |
| 210 | レコード抽出部 |
| 230 | 遷移ベクトル抽出部 |
| 500 | 履歴情報記憶部 |

- 5 1 0 データセット
- 5 2 1 前提レコード分
- 5 2 2 結論レコード分
- 5 3 0 抽出レコード群
- 5 3 1 抽出前提レコード群
- 5 3 2 抽出結論レコード群
- 5 4 0 共通部分レコード群
- 5 4 1 共通部分前提レコード群
- 5 4 2 共通部分結論レコード群
- 5 5 0 結論ソートレコード群
- 5 5 1 結論ソート前提レコード群
- 5 5 2 結論ソート結論レコード群
- 5 6 2 匿名グループ結論レコード群
- 5 7 0 残レコード
- 6 0 0 匿名化情報記憶部
- 6 1 1 前提匿名グループデータセット
- 6 1 2 結論匿名グループデータセット
- 7 0 0 コンピュータ
- 7 0 1 C P U
- 7 0 2 記憶部
- 7 0 3 記憶装置
- 7 0 4 入力部
- 7 0 5 出力部
- 7 0 6 通信部
- 7 0 7 記録媒体
- 5 3 2 1 結論レコード

請求の範囲

[請求項1] 固有識別子及び少なくとも1つの第1の属性を含む第1のレコードと、前記固有識別子と同一の固有識別子及び少なくとも1つの第2の属性を含む第2のレコードと、の組が複数件含まれるデータセットの中から、複数の前記第2のレコードを含む第2のレコード群において第2の1-多様性を充足可能であること、前記第2のレコード群に含まれる第2のレコードと組を成す前記第1のレコードから成る前記第1のレコード群において第1の1-多様性を充足可能であること、及び前記第1のレコードと前記第2のレコードとの間に存在する対応関係の抽象度に基づいて、複数の前記第2のレコードを抽出するレコード抽出手段と、

前記レコード抽出手段によって抽出された前記第2のレコードからなる匿名グループデータセットを、前記匿名グループデータセットにおいて前記第2の1-多様性を充足可能であり、かつ前記匿名グループデータセットに含まれる第2のレコードと組を成す前記第1のレコードからなる第1のレコード群において前記第1の1-多様性を充足可能であるように、生成し、出力する匿名グループ生成手段と、を備える

情報処理装置。

[請求項2] 前記匿名グループ生成手段は、更に、前記匿名グループデータセット及び前記匿名グループデータセットに含まれる第2のレコードのそれぞれと組を成す複数の前記第1のレコードが匿名化された前提匿名グループデータセットに対し、前記匿名グループデータセットに含まれる第2のレコードと前記前提匿名グループデータセットに含まれる第1のレコードとの前記対応関係を示す情報を付与して出力する

ことを特徴とする請求項1記載の情報処理装置。

[請求項3] 前記レコード抽出手段は、
前記第1のレコードに含まれる前記第1の属性の属性値毎の、前記

第2のレコードに含まれる第2の属性の各第2の属性値が、前記第1のレコードと前記組を成す前記第2のレコードに出現する頻度を要素とする遷移ベクトルを生成し、

2つの前記遷移ベクトルのそれぞれに対応する前記第2のレコードのそれぞれ同士で同一である前記第2の属性の第2の属性値の数が、前記第2の1-多様性の種類数未満である前記遷移ベクトル間の類似度を最低値の0として、前記遷移ベクトル間の類似度を算出し、

前記類似度が相対的に大きい順の、前記第1の1-多様性の種類数の前記遷移ベクトルのそれぞれに対応する前記第1の属性値を含む第1レコードと組を成す前記第2のレコードを、前記抽象度が相対的に小さい前記第2のレコードとして抽出する、

ことを特徴とする請求項1または2記載の情報処理装置。

[請求項4] 複数の前記遷移ベクトルについての前記類似度の算出対象を示す算出対象情報を生成し、前記算出対象情報を出力する遷移ベクトル抽出手段を更に含み、

前記レコード抽出手段は、前記生成した遷移ベクトルを前記遷移ベクトル抽出手段に出力し、前記遷移ベクトル抽出手段から前記算出対象情報を取得する

ことを特徴とする請求項3記載の情報処理装置。

[請求項5] 前記レコード抽出手段は、前記抽出した第1のレコードに対応する前記遷移ベクトルを除いた、前記生成した遷移ベクトルを遷移ベクトル抽出手段に出力する

ことを特徴とする請求項4記載の情報処理装置。

[請求項6] 前記匿名グループ生成手段は、前記匿名グループデータセットに含まれる第2のレコードの第2の属性の属性値と、匿名化された前記第1のレコード群に含まれる第1のレコードの第1の属性の属性値との間の前記対応関係の種類が増加しないように、前記匿名グループデータセットを生成する、

ことを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の情報処理装置。

[請求項7] 前記匿名グループ生成手段は、更に、前記匿名グループデータセットに前記対応関係の抽象化が発生しないように追加可能な、前記匿名グループデータセットに含まれていない、前記第 2 のレコードを前記匿名グループデータセットに追加する、

ことを特徴とする請求項 6 記載の情報処理装置。

[請求項8] 前記匿名グループ生成手段は、更に、前記匿名グループデータセットに含まれていない前記第 2 のレコードから、前記第 2 の 1 ー多様性を充足する匿名化が可能な前記第 2 のレコードの組であって、前記第 2 の 1 ー多様性を充足する匿名化が可能な前記第 2 のレコードと組を成す前記第 1 のレコードの組において前記第 1 の 1 ー多様性を充足可能である、前記第 2 のレコードの組を抽出し、前記匿名グループデータセットに追加する

ことを特徴とする請求項 6 または 7 記載の情報処理装置。

[請求項9] コンピュータが、

固有識別子及び少なくとも 1 つの第 1 の属性を含む第 1 のレコードと、前記固有識別子と同一の固有識別子及び少なくとも 1 つの第 2 の属性を含む第 2 のレコードと、の組が複数件含まれるデータセットの中から、前記第 2 のレコードからなる第 2 のレコード群において第 2 の 1 ー多様性を充足可能であること、前記第 2 のレコード群に含まれる第 2 のレコードと組を成す前記第 1 のレコードから成る前記第 1 のレコード群において第 1 の 1 ー多様性を充足可能であること、及び前記第 1 のレコードと前記第 2 のレコードとの間に存在する対応関係の抽象度に基づいて、複数の前記第 2 のレコードを抽出し、

前記抽出された前記第 2 のレコードからなる匿名グループデータセットを、前記匿名グループデータセットにおいて前記第 2 の 1 ー多様性を充足可能であり、かつ前記匿名グループデータセットに含まれる

第2のレコードと組を成す前記第1のレコードからなる第1のレコード群において前記第1の1ー多様性を充足可能であるように、生成し、出力する

匿名化方法。

[請求項10]

前記第2のレコードの抽出は、

前記第1のレコードに含まれる前記第1の属性の属性値毎の、前記第2のレコードに含まれる第2の属性の第2の各属性値が、前記第1のレコードと前記組を成す前記第2のレコードに出現する頻度を要素とする遷移ベクトルを生成し、

2つの前記遷移ベクトルそれぞれに対応する第2のレコードのそれぞれ同士で同一である、前記第2の属性の第2の属性値の数が、前記第2の1ー多様性の種類数未満である前記遷移ベクトル間の類似度を最低値の0として、前記遷移ベクトル間の類似度を算出し、

前記類似度が相対的に大きい順の、前記第1の1ー多様性の種類数の前記遷移ベクトルそれぞれに対応する前記第1の属性値を含む第1レコードと組を成す前記第2のレコードとを、前記抽象度が相対的に小さい前記第2のレコードとして抽出する、

ことを特徴とする請求項9記載の匿名化方法。

[請求項11]

前記コンピュータが、更に、複数の前記遷移ベクトルについての前記類似度の算出対象を示す算出対象情報を生成し、前記算出対象情報を出力し、

前記第2のレコードの抽出において、前記生成した遷移ベクトルに対応する前記算出対象情報に基づいて、前記遷移ベクトル間の類似度を算出する

ことを特徴とする請求項10記載の匿名化方法。

[請求項12]

固有識別子及び少なくとも1つの第1の属性を含む第1のレコードと、前記固有識別子と同一の固有識別子及び少なくとも1つの第2の属性を含む第2のレコードと、の組が複数件含まれるデータセットの

中から、前記第2のレコードからなる第2のレコード群において第2の1ー多様性を充足可能であること、前記第2のレコード群に含まれる第2のレコードと組を成す前記第1のレコードから成る前記第1のレコード群において第1の1ー多様性を充足可能であること、及び前記第1のレコードと前記第2のレコードとの間に存在する対応関係の抽象度に基づいて、複数の前記第2のレコードを抽出する処理と、

前記抽出された前記第2のレコードからなる匿名グループデータセットを、前記匿名グループデータセットにおいて前記第2の1ー多様性を充足可能であり、かつ前記匿名グループデータセットに含まれる第2のレコードと組を成す前記第1のレコードからなる第1のレコード群において前記第1の1ー多様性を充足可能であるように、生成し、出力する処理と、をコンピュータに実行させるための

プログラムを記録したコンピュータ読み取り可能不揮発性記録媒体

。

[請求項13]

前記第2のレコードを抽出する処理において、

前記第1のレコードに含まれる前記第1の属性の属性値毎の、前記第2のレコードに含まれる第2の属性の各第2の属性値が、前記第1のレコードと前記組を成す前記第2のレコード出現する頻度を要素とする遷移ベクトルを生成し、

2つの前記遷移ベクトルそれぞれに対応する第2のレコードのそれぞれ同士で同一である、前記第2の属性の第2の属性値の数が、前記第2の1ー多様性の種類数未満である前記遷移ベクトル間の類似度を最低値の0として、前記遷移ベクトル間の類似度を算出し、

前記類似度が相対的に大きい順の、前記第1の1ー多様性の種類数の前記遷移ベクトルそれぞれに対応する前記第1の属性値を含む第1レコードと組を成す前記第2のレコードとを、前記抽象度が相対的に小さい前記第2のレコードとして抽出する、処理を前記コンピュータに実行させる

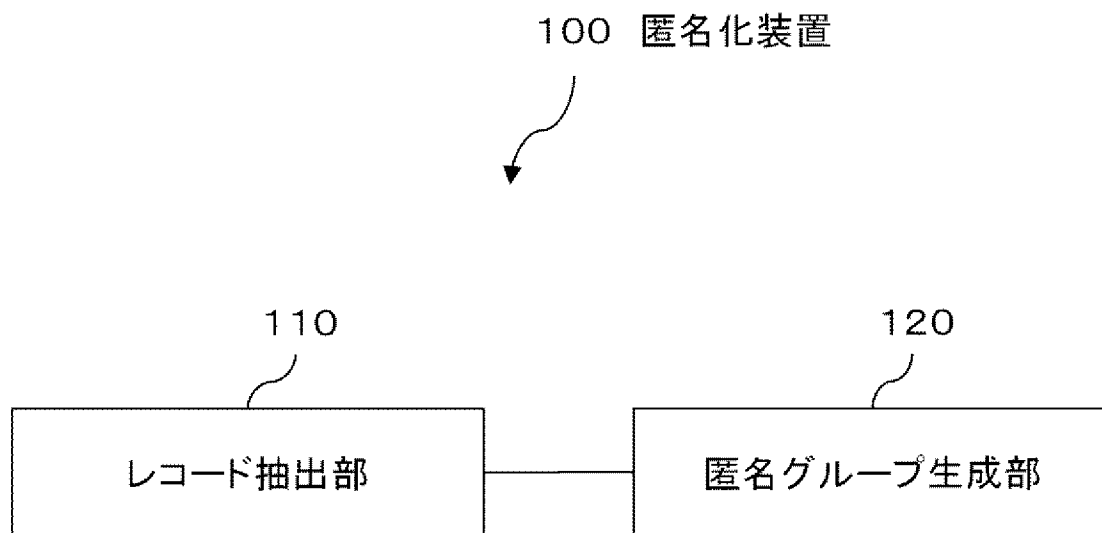
前記プログラムを記録した請求項 1 2 記載のコンピュータ読み取り可能不揮発性記録媒体。

[請求項14] 複数の前記遷移ベクトルについての前記類似度の算出対象を示す算出対象情報を生成し、前記算出対象情報を出力する処理を、更に、前記コンピュータに実行させ、

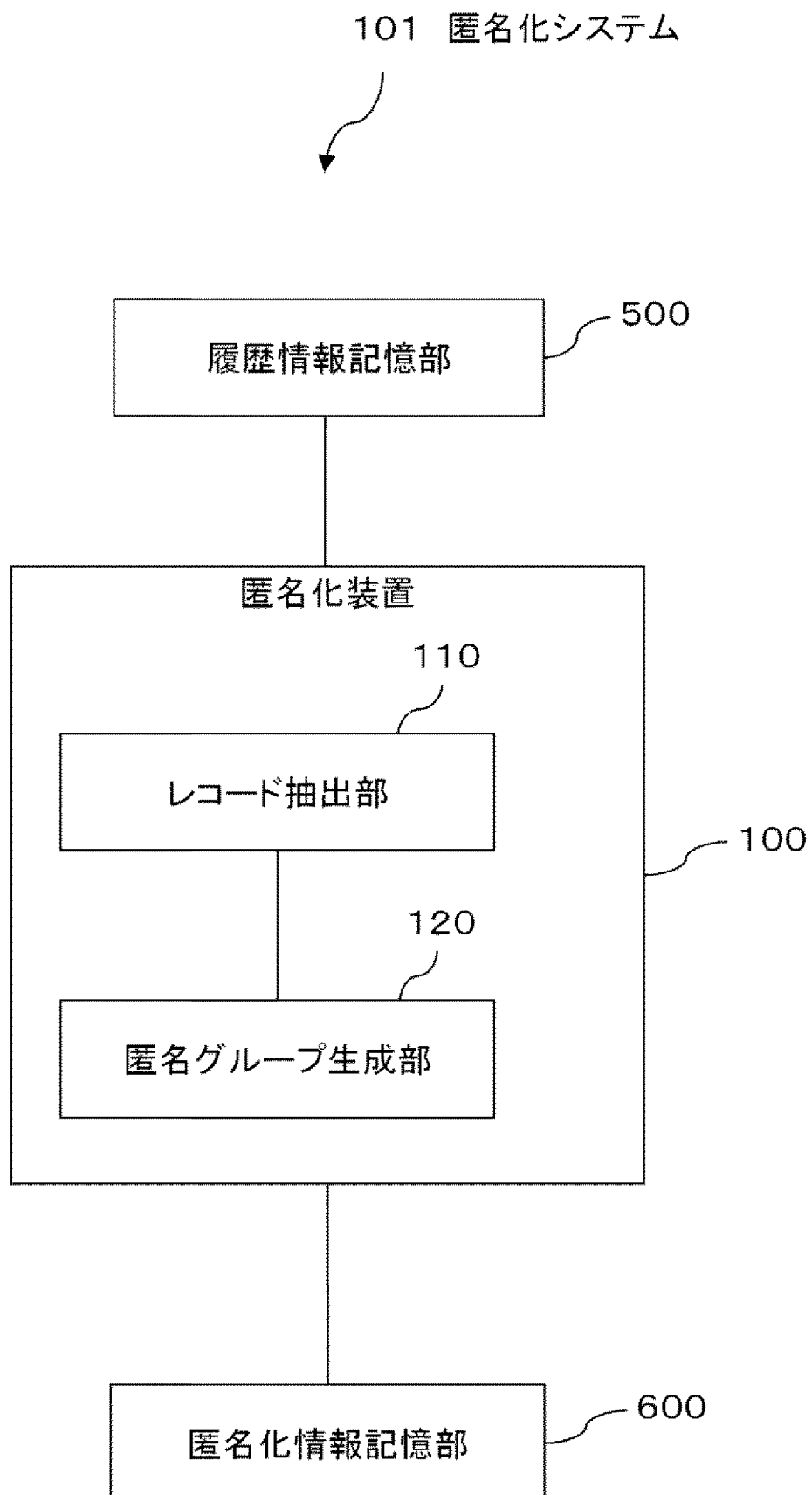
前記第 2 のレコードの抽出において、前記生成した遷移ベクトルに対応する前記算出対象情報に基づいて、前記遷移ベクトル間の類似度を算出する、処理を前記コンピュータに実行させる

前記プログラムを記録した請求項 1 3 記載のプログラムを記録した不揮発性記録媒体。

[図1]



[図2]



[図3]

510 データセット



固有識別子	診療月	年齢	病名
1	4月	21	U
2	4月	22	V
3	4月	23	V
4	4月	24	W
5	4月	25	Y
6	4月	26	Y
7	4月	27	X
8	4月	28	X
9	4月	29	W
10	4月	30	V
.	.	.	.
.	.	.	.
.	.	.	.
34	4月	54	Y
35	4月	55	Z
36	4月	56	W
37	4月	57	W
38	4月	58	U
39	4月	59	U
1	5月	21	A
2	5月	22	A
3	5月	23	D
4	5月	24	B
.	.	.	.
.	.	.	.
.	.	.	.
29	5月	49	C
30	5月	50	B
31	5月	51	C
32	5月	52	A
33	5月	53	D
34	5月	54	D
35	5月	55	F
36	5月	56	C
37	5月	57	A
38	5月	58	C
39	5月	59	A

[図4]

521 前提レコード分



固有識別子	診療月	年齢	病名
1	4月	21	U
13	4月	33	U
27	4月	47	U
39	4月	59	U
14	4月	34	U
26	4月	46	U
28	4月	48	U
29	4月	49	U
38	4月	58	U
11	4月	31	U
12	4月	32	U
2	4月	22	V
25	4月	45	V
10	4月	30	V
15	4月	35	V
16	4月	36	V
30	4月	50	V
24	4月	44	V
31	4月	51	V
3	4月	23	V
32	4月	52	W
37	4月	57	W
4	4月	24	W
22	4月	42	W
23	4月	43	W
9	4月	29	W
17	4月	37	W
36	4月	56	W
33	4月	53	W
8	4月	28	X
18	4月	38	X
19	4月	39	X
7	4月	27	X
21	4月	41	X
5	4月	25	Y
34	4月	54	Y
6	4月	26	Y
20	4月	40	Z
35	4月	55	Z

[図5]

522 結論レコード分



固有識別子	診療月	年齢	病名
1	5月	21	A
13	5月	33	A
27	5月	47	A
39	5月	59	A
14	5月	34	B
26	5月	46	B
28	5月	48	B
29	5月	49	C
38	5月	58	C
11	5月	31	D
12	5月	32	D
2	5月	22	A
25	5月	45	A
10	5月	30	B
15	5月	35	B
16	5月	36	B
30	5月	50	B
24	5月	44	C
31	5月	51	C
3	5月	23	D
32	5月	52	A
37	5月	57	A
4	5月	24	B
22	5月	42	B
23	5月	43	B
9	5月	29	C
17	5月	37	C
36	5月	56	C
33	5月	53	D
8	5月	28	A
18	5月	38	B
19	5月	39	E
7	5月	27	D
21	5月	41	E
5	5月	25	D
34	5月	54	D
6	5月	26	E
20	5月	40	F
35	5月	55	F

[図6]

611 前提匿名グループデータセット



固有 識別子	グループ識 別子	関連 識別子	診療月	年齢	病名
1	101	203	4月	21-52	U
2	101	203	4月	21-52	V
32	101	203	4月	21-52	W
13	102	201	4月	33-57	U
25	102	201	4月	33-57	V
37	102	201	4月	33-57	W
14	103	204	4月	24-34	U
10	103	204	4月	24-34	V
4	103	204	4月	24-34	W
26	104	202	4月	35-46	U
15	104	202	4月	35-46	V
22	104	202	4月	35-46	W
28	105	201	4月	36-48	U
16	105	201	4月	36-48	V
23	105	201	4月	36-48	W
29	106	203	4月	29-49	U
24	106	203	4月	29-49	V
9	106	203	4月	29-49	W
38	107	202	4月	37-58	U
31	107	202	4月	37-58	V
17	107	202	4月	37-58	W
11	108	204	4月	23-31	U
3	108	204	4月	23-31	V
33	108	204	4月	23-31	W

[図7]

612 結論匿名グループデータセット



固有 識別子	グループ識 別子	関連 識別子	診療月	年齢	病名
13	201	000	5月	33-57	A
25	201	000	5月	33-57	A
37	201	000	5月	33-57	A
28	201	000	5月	33-57	B
16	201	000	5月	33-57	B
23	201	000	5月	33-57	B
26	202	000	5月	46-58	B
15	202	000	5月	46-58	B
22	202	000	5月	46-58	B
38	202	000	5月	46-58	C
31	202	000	5月	46-58	C
17	202	000	5月	46-58	C
1	203	000	5月	21-52	A
2	203	000	5月	21-52	A
32	203	000	5月	21-52	A
29	203	000	5月	21-52	C
24	203	000	5月	21-52	C
9	203	000	5月	21-52	C
14	204	000	5月	23-53	B
10	204	000	5月	23-53	B
4	204	000	5月	23-53	B
11	204	000	5月	23-53	D
3	204	000	5月	23-53	D
33	204	000	5月	23-53	D

[図8]

530 抽出レコード群

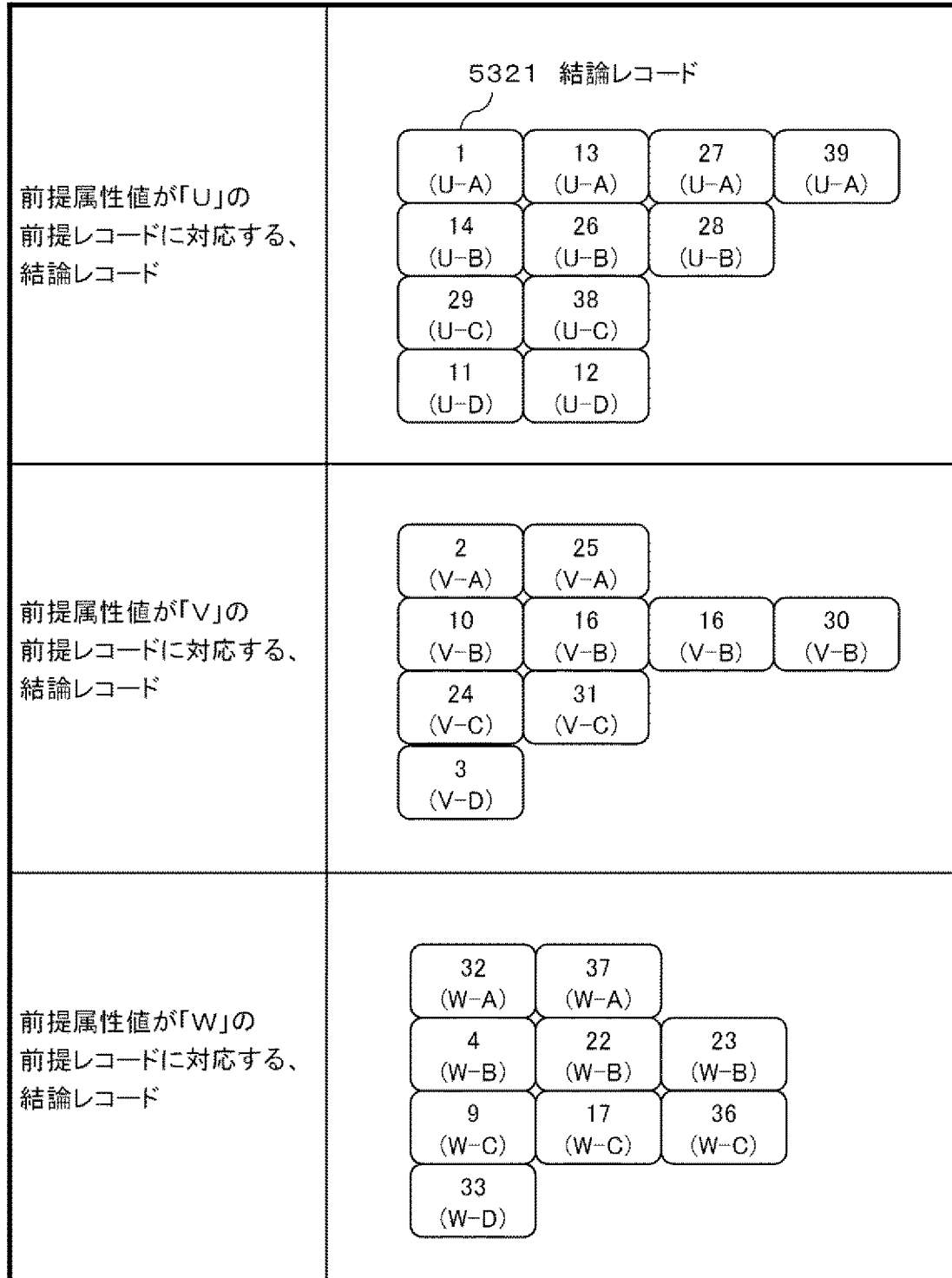
531 抽出前提レコード群

532 抽出結論レコード群

固有 識別子	診療月	年齢	病名	固有 識別子	診療月	年齢	病名
1	4月	21	U	1	5月	21	A
13	4月	33	U	13	5月	33	A
27	4月	47	U	27	5月	47	A
39	4月	59	U	39	5月	59	A
14	4月	34	U	14	5月	34	B
26	4月	46	U	26	5月	46	B
28	4月	48	U	28	5月	48	B
29	4月	49	U	29	5月	49	C
38	4月	58	U	38	5月	58	C
11	4月	31	U	11	5月	31	D
12	4月	32	U	12	5月	32	D
2	4月	22	V	2	5月	22	A
25	4月	45	V	25	5月	45	A
10	4月	30	V	10	5月	30	B
15	4月	35	V	15	5月	35	B
16	4月	36	V	16	5月	36	B
30	4月	50	V	30	5月	50	B
24	4月	44	V	24	5月	44	C
31	4月	51	V	31	5月	51	C
3	4月	23	V	3	5月	23	D
32	4月	52	W	32	5月	52	A
37	4月	57	W	37	5月	57	A
4	4月	24	W	4	5月	24	B
22	4月	42	W	22	5月	42	B
23	4月	43	W	23	5月	43	B
9	4月	29	W	9	5月	29	C
17	4月	37	W	17	5月	37	C
36	4月	56	W	36	5月	56	C
33	4月	53	W	33	5月	53	D

[図9]

532 抽出結論レコード群



[図10]

540 共通部分レコード群



541 共通部分前提レコード群

542 共通部分結論レコード群

固有 識別子	診療月	年齢	病名	固有 識別子	診療月	年齢	病名
1	4月	21	U	1	5月	21	A
13	4月	33	U	13	5月	33	A
14	4月	34	U	14	5月	34	B
26	4月	46	U	26	5月	46	B
28	4月	48	U	28	5月	48	B
29	4月	49	U	29	5月	49	C
38	4月	58	U	38	5月	58	C
11	4月	31	U	11	5月	31	D
2	4月	22	V	2	5月	22	A
25	4月	45	V	25	5月	45	A
10	4月	30	V	10	5月	30	B
15	4月	35	V	15	5月	35	B
16	4月	36	V	16	5月	36	B
24	4月	44	V	24	5月	44	C
31	4月	51	V	31	5月	51	C
3	4月	23	V	3	5月	23	D
32	4月	52	W	32	5月	52	A
37	4月	57	W	37	5月	57	A
4	4月	24	W	4	5月	24	B
22	4月	42	W	22	5月	42	B
23	4月	43	W	23	5月	43	B
9	4月	29	W	9	5月	29	C
17	4月	37	W	17	5月	37	C
33	4月	53	W	33	5月	53	D

[図11]

542 共通部分結論レコード群



前提属性値が「U」の 前提レコードに対応する、 結論レコード群 の共通部分	<table><tr><td>1 (U-A)</td><td>13 (U-A)</td><td></td></tr><tr><td>14 (U-B)</td><td>26 (U-B)</td><td>28 (U-B)</td></tr><tr><td>29 (U-C)</td><td>38 (U-C)</td><td></td></tr><tr><td>11 (U-D)</td><td></td><td></td></tr></table>	1 (U-A)	13 (U-A)		14 (U-B)	26 (U-B)	28 (U-B)	29 (U-C)	38 (U-C)		11 (U-D)		
1 (U-A)	13 (U-A)												
14 (U-B)	26 (U-B)	28 (U-B)											
29 (U-C)	38 (U-C)												
11 (U-D)													
前提属性値が「V」の 前提レコードに対応する、 結論レコード群 の共通部分	<table><tr><td>2 (V-A)</td><td>25 (V-A)</td><td></td></tr><tr><td>10 (V-B)</td><td>16 (V-B)</td><td>16 (V-B)</td></tr><tr><td>24 (V-C)</td><td>31 (V-C)</td><td></td></tr><tr><td>3 (V-D)</td><td></td><td></td></tr></table>	2 (V-A)	25 (V-A)		10 (V-B)	16 (V-B)	16 (V-B)	24 (V-C)	31 (V-C)		3 (V-D)		
2 (V-A)	25 (V-A)												
10 (V-B)	16 (V-B)	16 (V-B)											
24 (V-C)	31 (V-C)												
3 (V-D)													
前提属性値が「W」の 前提レコードに対応する、 結論レコード群 の共通部分	<table><tr><td>32 (W-A)</td><td>37 (W-A)</td><td></td></tr><tr><td>4 (W-B)</td><td>22 (W-B)</td><td>23 (W-B)</td></tr><tr><td>9 (W-C)</td><td>17 (W-C)</td><td></td></tr><tr><td>33 (W-D)</td><td></td><td></td></tr></table>	32 (W-A)	37 (W-A)		4 (W-B)	22 (W-B)	23 (W-B)	9 (W-C)	17 (W-C)		33 (W-D)		
32 (W-A)	37 (W-A)												
4 (W-B)	22 (W-B)	23 (W-B)											
9 (W-C)	17 (W-C)												
33 (W-D)													

[図12]

550 結論ソートレコード群

551 結論ソート前提レコード群

552 結論ソート結論レコード群

固有 識別子	診療月	年齢	病名	固有 識別子	診療月	年齢	病名
1	4月	21	U	1	5月	21	A
13	4月	33	U	13	5月	33	A
2	4月	22	V	2	5月	22	A
25	4月	45	V	25	5月	45	A
32	4月	52	W	32	5月	52	A
37	4月	57	W	37	5月	57	A
14	4月	34	U	14	5月	34	B
26	4月	46	U	26	5月	46	B
28	4月	48	U	28	5月	48	B
10	4月	30	V	10	5月	30	B
15	4月	35	V	15	5月	35	B
16	4月	36	V	16	5月	36	B
4	4月	24	W	4	5月	24	B
22	4月	42	W	22	5月	42	B
23	4月	43	W	23	5月	43	B
29	4月	49	U	29	5月	49	C
38	4月	58	U	38	5月	58	C
24	4月	44	V	24	5月	44	C
31	4月	51	V	31	5月	51	C
9	4月	29	W	9	5月	29	C
17	4月	37	W	17	5月	37	C
11	4月	31	U	11	5月	31	D
3	4月	23	V	3	5月	23	D
33	4月	53	W	33	5月	53	D

[図13]

552 結論ソート結論レコード群



結論属性値が「A」の 結論レコード	<table><tr><td>1 (U-A)</td><td>2 (V-A)</td><td>32 (W-A)</td></tr><tr><td>13 (U-A)</td><td>25 (V-A)</td><td>37 (W-A)</td></tr></table>	1 (U-A)	2 (V-A)	32 (W-A)	13 (U-A)	25 (V-A)	37 (W-A)			
1 (U-A)	2 (V-A)	32 (W-A)								
13 (U-A)	25 (V-A)	37 (W-A)								
結論属性値が「B」の 結論レコード	<table><tr><td>14 (U-B)</td><td>10 (V-B)</td><td>4 (W-B)</td></tr><tr><td>26 (U-B)</td><td>16 (V-B)</td><td>22 (W-B)</td></tr><tr><td>28 (U-B)</td><td>16 (V-B)</td><td>23 (W-B)</td></tr></table>	14 (U-B)	10 (V-B)	4 (W-B)	26 (U-B)	16 (V-B)	22 (W-B)	28 (U-B)	16 (V-B)	23 (W-B)
14 (U-B)	10 (V-B)	4 (W-B)								
26 (U-B)	16 (V-B)	22 (W-B)								
28 (U-B)	16 (V-B)	23 (W-B)								
結論属性値が「C」の 結論レコード	<table><tr><td>29 (U-C)</td><td>24 (V-C)</td><td>9 (W-C)</td></tr><tr><td>38 (U-C)</td><td>31 (V-C)</td><td>17 (W-C)</td></tr></table>	29 (U-C)	24 (V-C)	9 (W-C)	38 (U-C)	31 (V-C)	17 (W-C)			
29 (U-C)	24 (V-C)	9 (W-C)								
38 (U-C)	31 (V-C)	17 (W-C)								
結論属性値が「D」の 結論レコード	<table><tr><td>11 (U-D)</td><td>3 (V-D)</td><td>33 (W-D)</td></tr></table>	11 (U-D)	3 (V-D)	33 (W-D)						
11 (U-D)	3 (V-D)	33 (W-D)								

[図14]

562 匿名グループ結論レコード群

前提レコード群				結論レコード群				
固有 識別子	診療月	年齢	病名	固有 識別子	グループ識 別子	診療月	年齢	病名
1	4月	21	U	1	203	5月	21	A
13	4月	33	U	13	201	5月	33	A
2	4月	22	V	2	203	5月	22	A
25	4月	45	V	25	201	5月	45	A
32	4月	52	W	32	203	5月	52	A
37	4月	57	W	37	201	5月	57	A
14	4月	34	U	14	204	5月	34	B
26	4月	46	U	26	202	5月	46	B
28	4月	48	U	28	201	5月	48	B
10	4月	30	V	10	204	5月	30	B
15	4月	35	V	15	202	5月	35	B
16	4月	36	V	16	201	5月	36	B
4	4月	24	W	4	204	5月	24	B
22	4月	42	W	22	202	5月	42	B
23	4月	43	W	23	201	5月	43	B
29	4月	49	U	29	203	5月	49	C
38	4月	58	U	38	202	5月	58	C
24	4月	44	V	24	203	5月	44	C
31	4月	51	V	31	202	5月	51	C
9	4月	29	W	9	203	5月	29	C
17	4月	37	W	17	202	5月	37	C
11	4月	31	U	11	204	5月	31	D
3	4月	23	V	3	204	5月	23	D
33	4月	53	W	33	204	5月	53	D

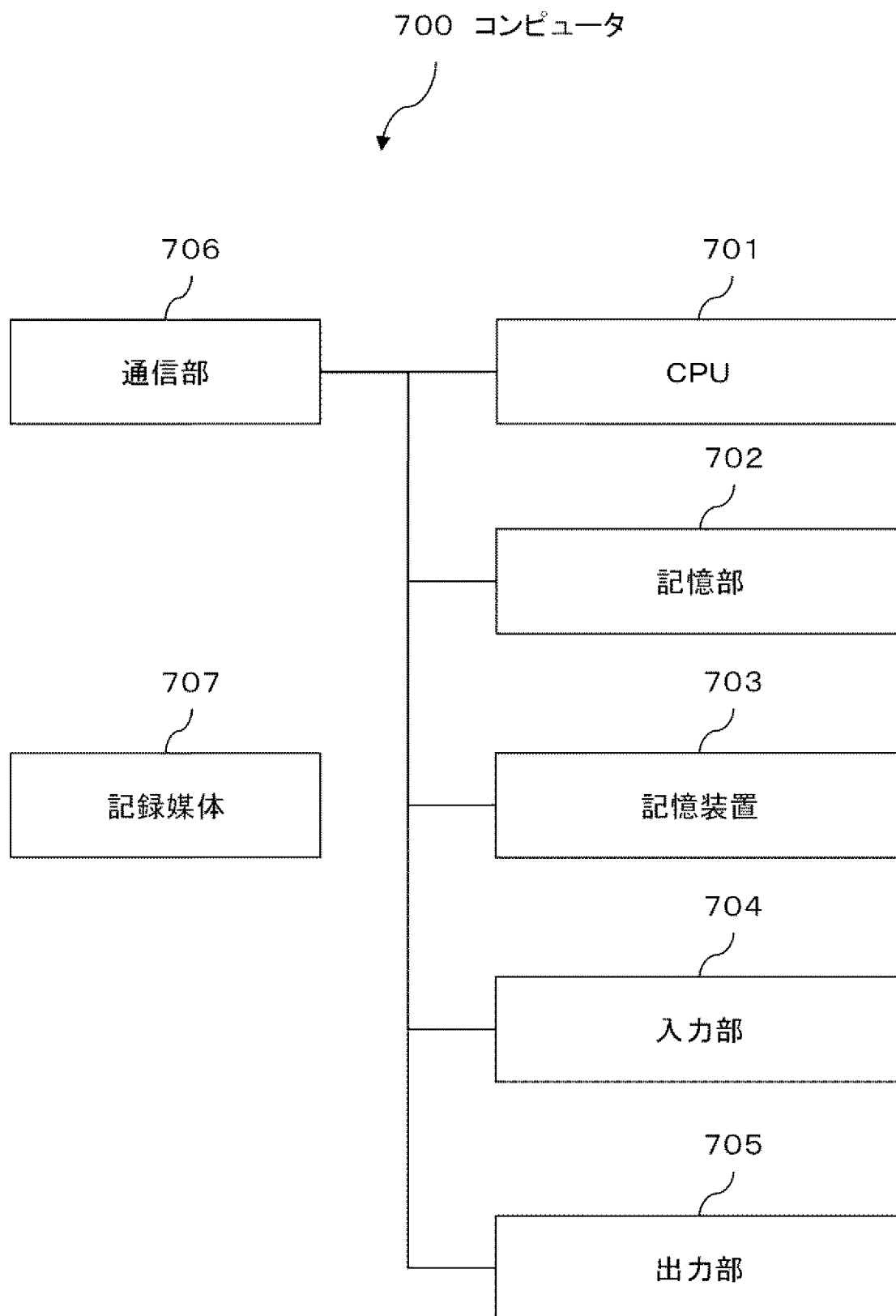
[図15]

562 匿名グループ結論レコード群

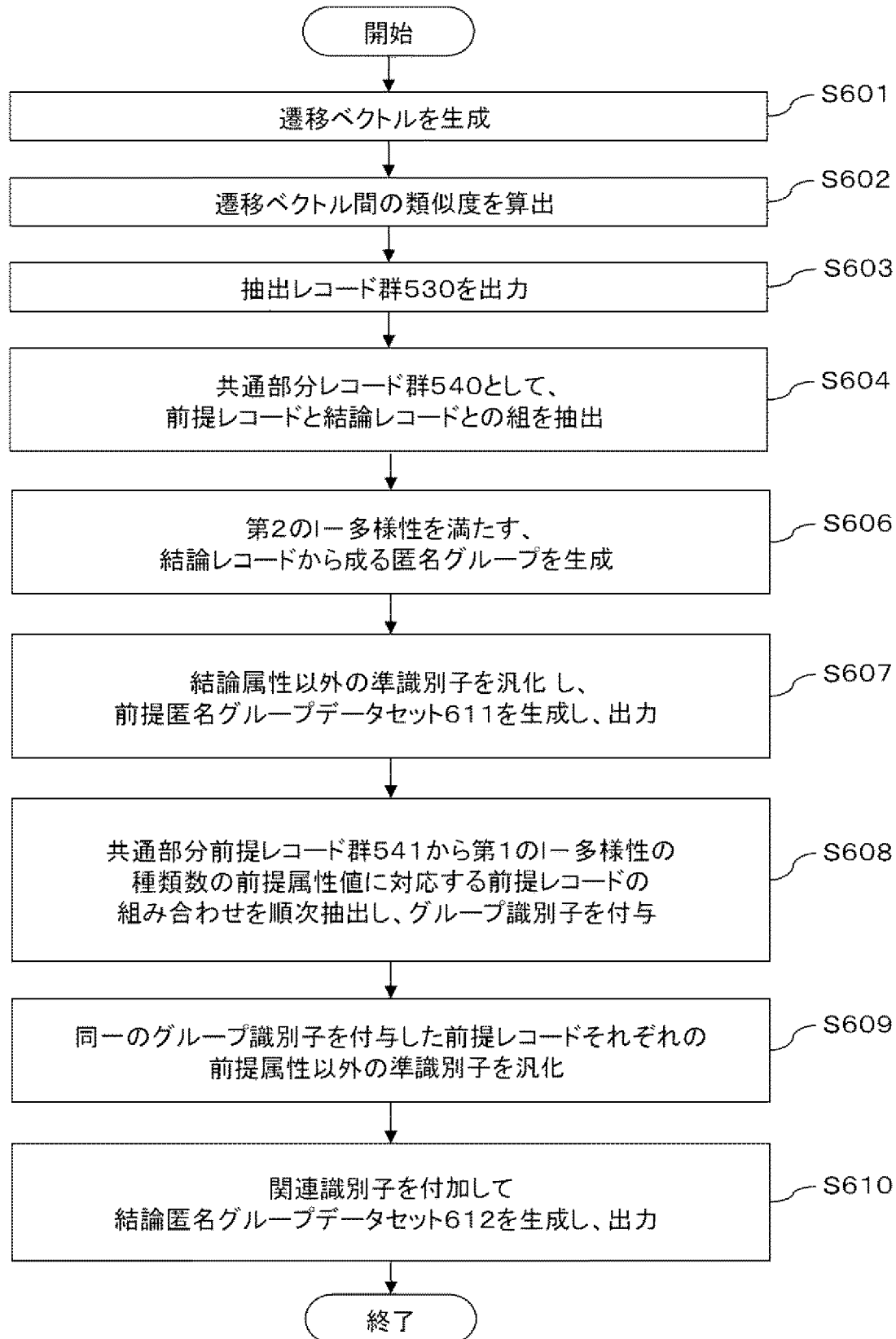


グループ識別子							
201	<table><tr><td>28 (U-B)</td><td>16 (V-B)</td><td>23 (W-B)</td></tr><tr><td>13 (U-A)</td><td>25 (V-A)</td><td>37 (W-A)</td></tr></table>	28 (U-B)	16 (V-B)	23 (W-B)	13 (U-A)	25 (V-A)	37 (W-A)
28 (U-B)	16 (V-B)	23 (W-B)					
13 (U-A)	25 (V-A)	37 (W-A)					
202	<table><tr><td>26 (U-B)</td><td>16 (V-B)</td><td>22 (W-B)</td></tr><tr><td>38 (U-C)</td><td>31 (V-C)</td><td>17 (W-C)</td></tr></table>	26 (U-B)	16 (V-B)	22 (W-B)	38 (U-C)	31 (V-C)	17 (W-C)
26 (U-B)	16 (V-B)	22 (W-B)					
38 (U-C)	31 (V-C)	17 (W-C)					
203	<table><tr><td>1 (U-A)</td><td>2 (V-A)</td><td>32 (W-A)</td></tr><tr><td>29 (U-C)</td><td>24 (V-C)</td><td>9 (W-C)</td></tr></table>	1 (U-A)	2 (V-A)	32 (W-A)	29 (U-C)	24 (V-C)	9 (W-C)
1 (U-A)	2 (V-A)	32 (W-A)					
29 (U-C)	24 (V-C)	9 (W-C)					
204	<table><tr><td>14 (U-B)</td><td>10 (V-B)</td><td>4 (W-B)</td></tr><tr><td>11 (U-D)</td><td>3 (V-D)</td><td>33 (W-D)</td></tr></table>	14 (U-B)	10 (V-B)	4 (W-B)	11 (U-D)	3 (V-D)	33 (W-D)
14 (U-B)	10 (V-B)	4 (W-B)					
11 (U-D)	3 (V-D)	33 (W-D)					

[図16]



[図17]



[図18]

570 残レコード



27 (U-A)	39 (U-A)	12 (U-D)		
30 (V-B)				
36 (W-C)				
8 (X-A)	18 (X-B)	7 (X-D)	19 (X-E)	21 (X-E)
5 (Y-D)	34 (Y-D)	6 (Y-E)		
20 (Z-F)	35 (Z-F)			

[図19]

28 (U-B)	16 (V-B)	23 (W-B)	18 (X-B)
13 (U-A)	25 (V-A)	37 (W-A)	8 (X-A)

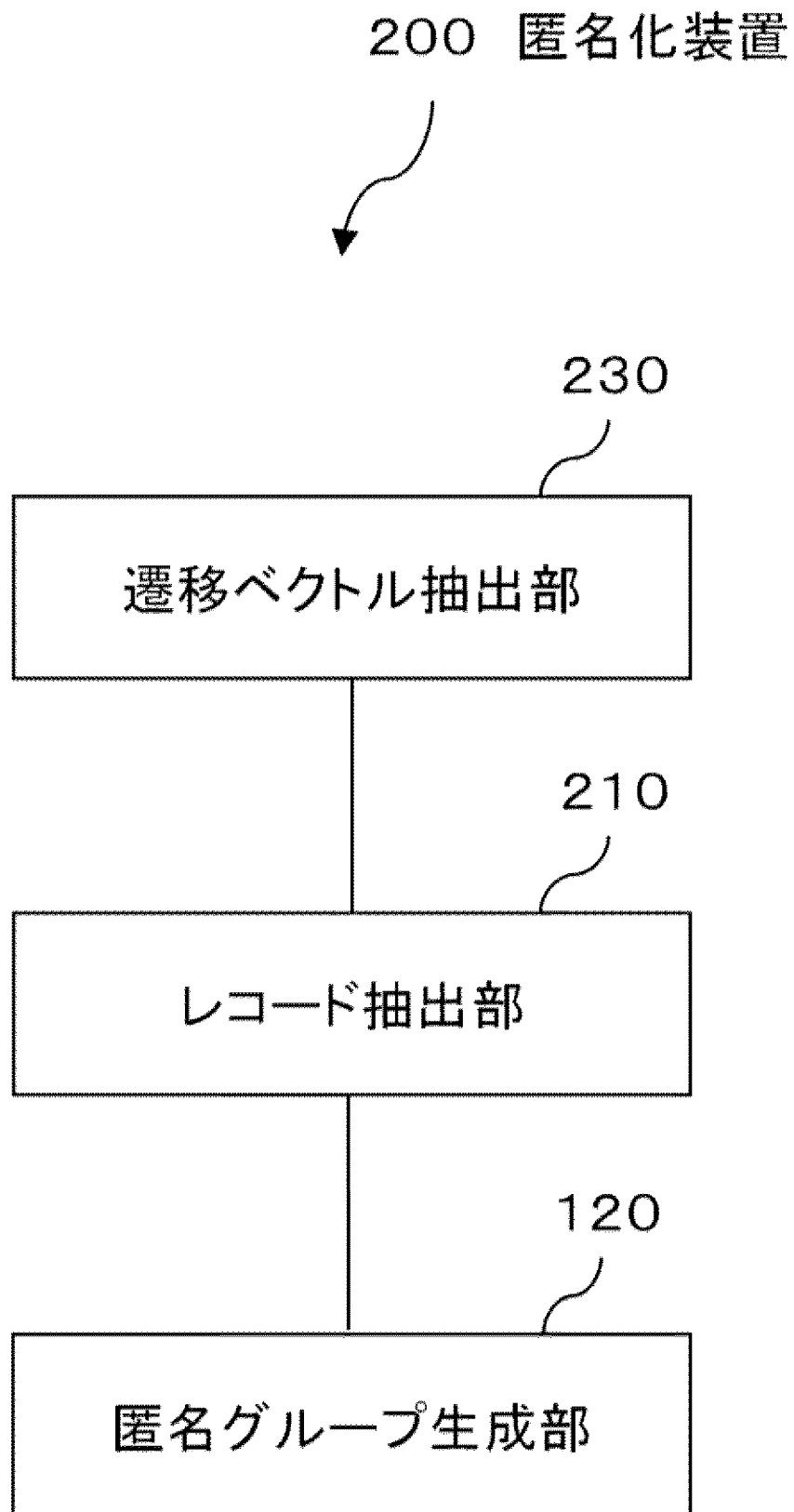
[図20]

28 (U-B)	91 (X-B)	92 (Y-B)
13 (U-A)	93 (X-A)	94 (Y-A)
12 (U-D)	7 (X-D)	5 (Y-D)

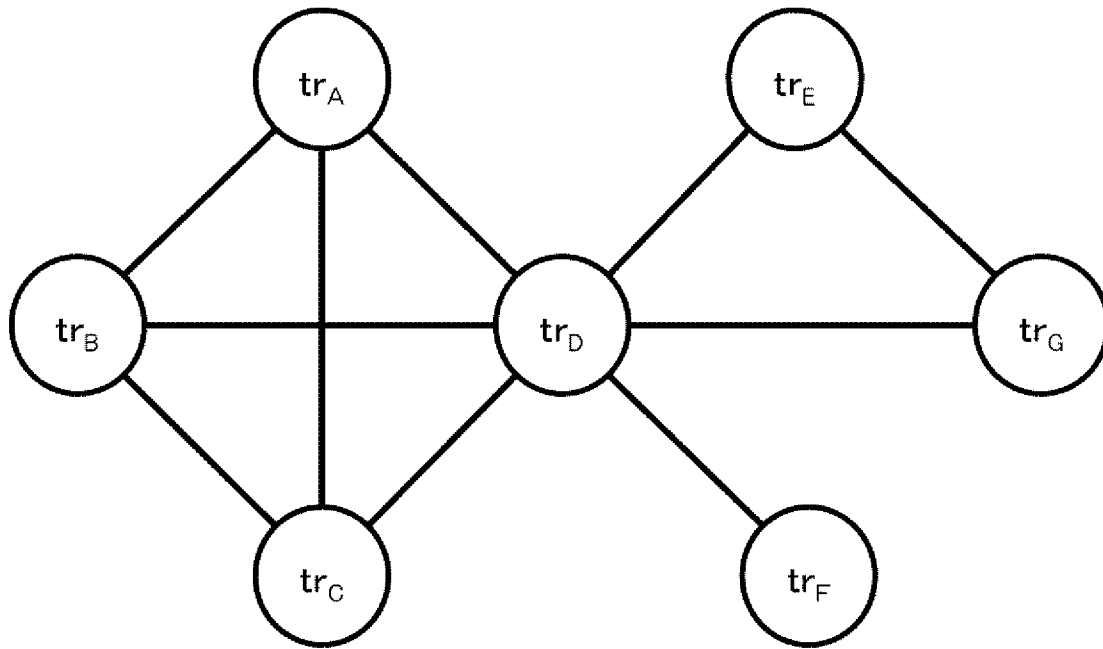
[図21]

27 (U-A)	39 (U-A)
7 (X-D)	19 (X-E)
34 (Y-D)	6 (Y-E)

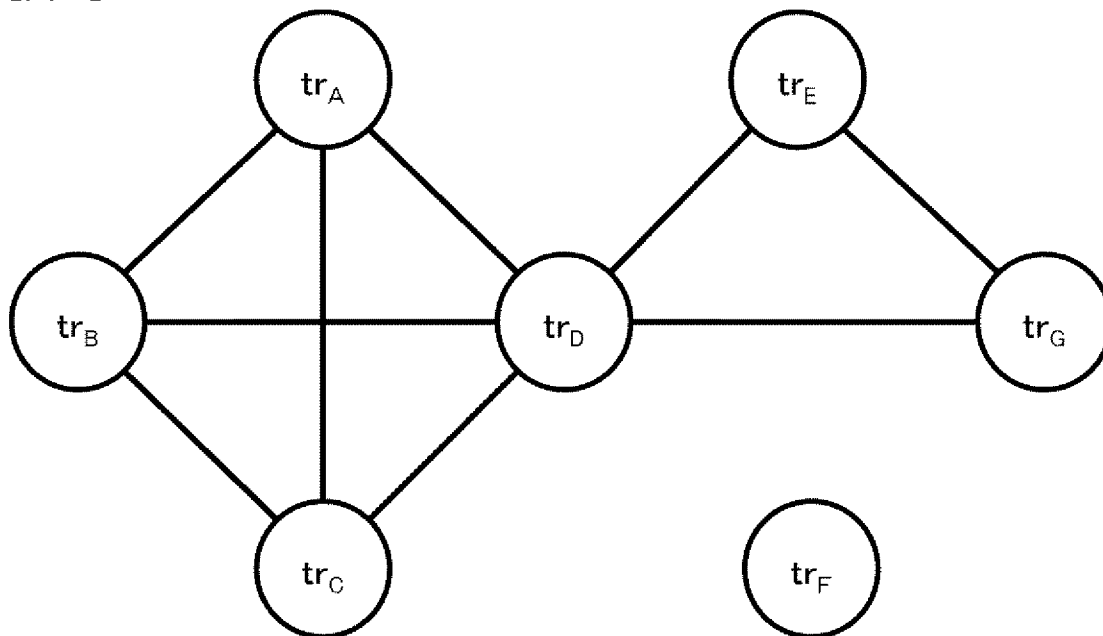
[図22]



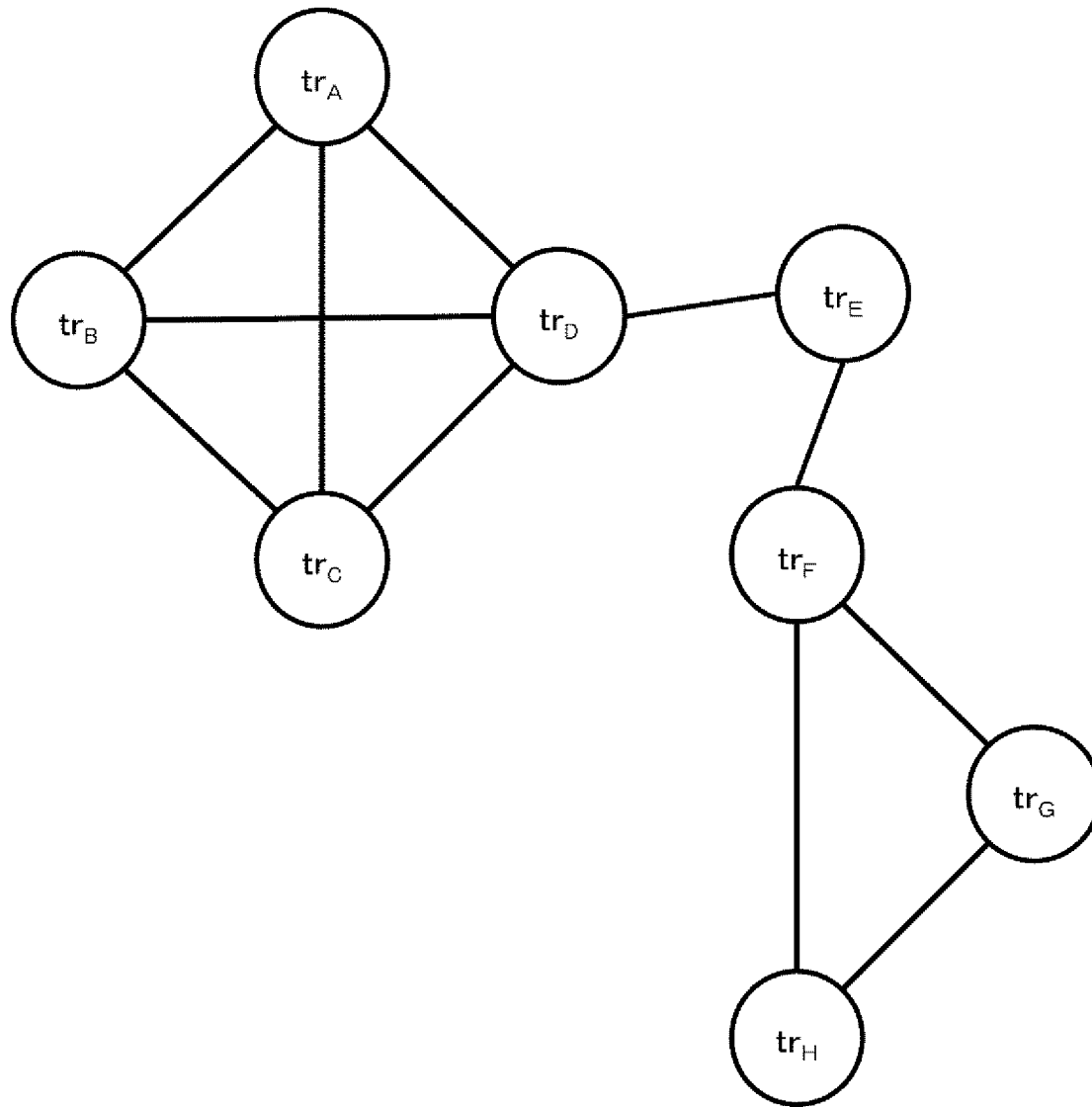
[図23]



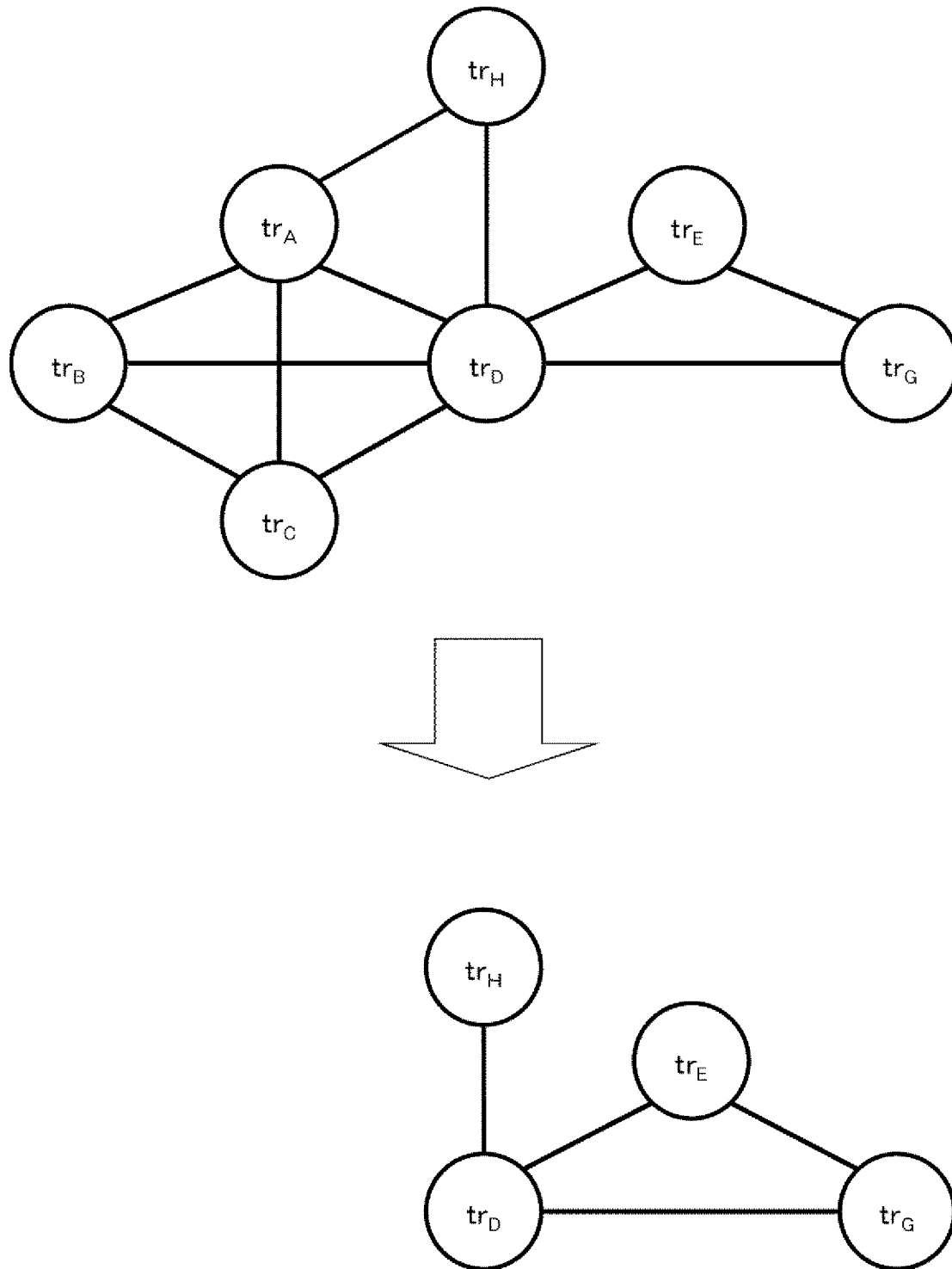
[図24]



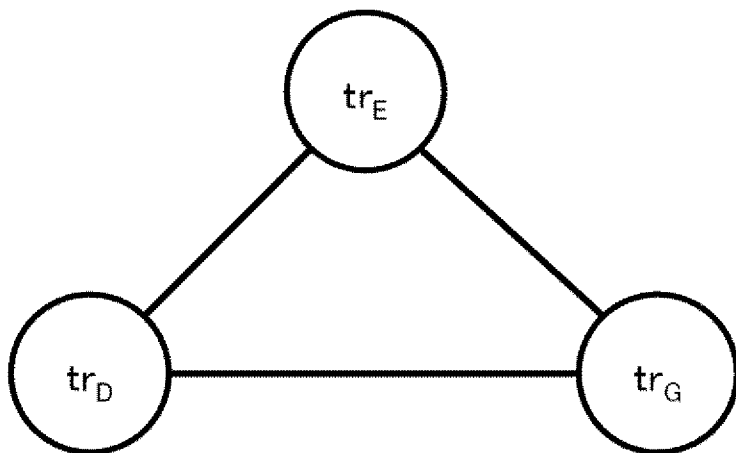
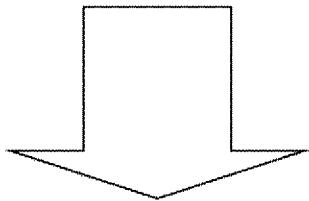
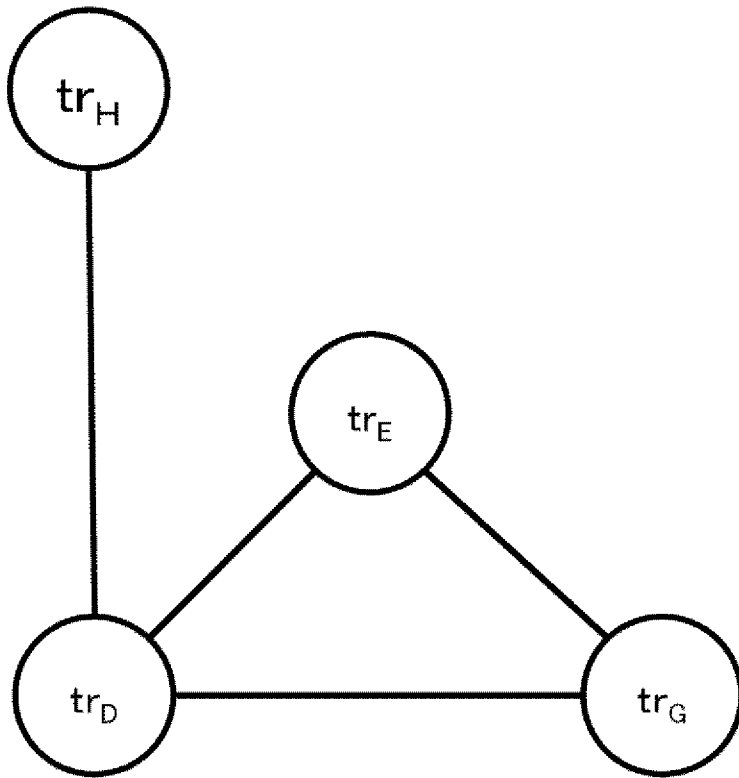
[図25]



[図26]



[図27]



[図28]

第1のレコード群				第2のレコード群			
固有 識別子	診療月	年齢	病名	固有 識別子	診療月	年齢	病名
1	4月	21	U	1	5月	21	A
2	4月	22	V	2	5月	22	A
3	4月	23	V	3	5月	23	D
4	4月	24	W	4	5月	24	B
5	4月	25	Y	5	5月	25	D
6	4月	26	Y	6	5月	26	E
7	4月	27	X	7	5月	27	D
8	4月	28	X	8	5月	28	A
9	4月	29	W	9	5月	29	C
10	4月	30	V	10	5月	30	B
11	4月	31	U	11	5月	31	D
12	4月	32	U	12	5月	32	D
13	4月	33	U	13	5月	33	A
14	4月	34	U	14	5月	34	B
15	4月	35	V	15	5月	35	B
16	4月	36	V	16	5月	36	B
17	4月	37	W	17	5月	37	C
18	4月	38	X	18	5月	38	B
19	4月	39	X	19	5月	39	E
20	4月	40	Z	20	5月	40	F
21	4月	41	X	21	5月	41	E
22	4月	42	W	22	5月	42	B
23	4月	43	W	23	5月	43	B
24	4月	44	V	24	5月	44	C
25	4月	45	V	25	5月	45	A
26	4月	46	U	26	5月	46	B
27	4月	47	U	27	5月	47	A
28	4月	48	U	28	5月	48	B
29	4月	49	U	29	5月	49	C
30	4月	50	V	30	5月	50	B
31	4月	51	V	31	5月	51	C
32	4月	52	W	32	5月	52	A
33	4月	53	W	33	5月	53	D
34	4月	54	Y	34	5月	54	D
35	4月	55	Z	35	5月	55	F
36	4月	56	W	36	5月	56	C
37	4月	57	W	37	5月	57	A
38	4月	58	U	38	5月	58	C
39	4月	59	U	39	5月	59	A

[図29]

I=3で匿名化された 第1のレコード群				I=2で匿名化された 第2のレコード群			
グループ 識別子	診療月	年齢	病名	グループ 識別子	診療月	年齢	病名
102	4月	21-24	U	102	5月	21-24	A
102	4月	21-24	V	102	5月	21-24	A
103	4月	23-28	V	103	5月	23-28	D
102	4月	21-24	W	102	5月	21-24	B
103	4月	23-28	Y	103	5月	23-28	D
101	4月	26-29	Y	101	5月	26-29	E
101	4月	26-29	X	101	5月	26-29	D
103	4月	23-28	X	103	5月	23-28	A
101	4月	26-29	W	101	5月	26-29	C
104	4月	30-37	V	104	5月	30-37	B
104	4月	30-37	U	104	5月	30-37	D
105	4月	32-38	U	105	5月	32-38	D
106	4月	33-39	U	106	5月	33-39	A
107	4月	34-41	U	107	5月	34-41	B
105	4月	32-38	V	105	5月	32-38	B
106	4月	33-39	V	106	5月	33-39	B
104	4月	30-37	W	104	5月	30-37	C
105	4月	32-38	X	105	5月	32-38	B
106	4月	33-39	X	106	5月	33-39	C
107	4月	34-41	Z	107	5月	34-41	F
107	4月	34-41	X	107	5月	34-41	E
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/005392

A. CLASSIFICATION OF SUBJECT MATTER

G06F21/60(2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F21/60

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Tsubasa TAKAHASHI, "Jikeiretsu Data ni Taisuru 1-Tayoka Hoshiki no Teian", Dai 4 Kai Forum on Data Engineering and Information Management Ronbunshu (Dai 10 Kai Database Society of Japan Nenji Taikai) [online], 30 August 2012 (30.08.2012) (received date), passages 3.6 to 4.2, Internet <URL:http://db-event.jpn.org/deim2012/proceedings/final-pdf/a1-1.pdf>	1-14
A	WO 2012/090628 A1 (NEC Corp.), 05 July 2012 (05.07.2012), abstract; claim 1 (Family: none)	1-14

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
02 December, 2013 (02.12.13)

Date of mailing of the international search report
10 December, 2013 (10.12.13)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/005392

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2012-159982 A (KDDI Corp.), 23 August 2012 (23.08.2012), abstract; claim 1 (Family: none)	1-14
A	Takao TAKENOUCHI, "Fukusu no Data Teikyo no Tameno Tokumeika", Computer Security Symposium 2013 Ronbunshu, vol.2013, no.4, Information Processing Society of Japan, 14 October 2013 (14.10.2013), pages 893 to 900	1-14

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F21/60(2013.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06F21/60

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 3 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 3 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 3 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	高橋 翼, 時系列データに対する 1 - 多様化方式の提案, 第 4 回データ工学と情報マネジメントに関するフォーラム論文集（第 1 0 回日本データベース学会年次大会）[o n l i n e], 2012.08.30（受入日）, 3. 6 節～4. 2 節, インターネット<URL:http://db-event.jp/jpn.org/deim2012/proceedings/final-pdf/a1-1.pdf>	1 - 1 4
A	W0 2012/090628 A1（日本電気株式会社）2012.07.05, 要約, 請求項 1（ファミリーなし）	1 - 1 4
A	JP 2012-159982 A（K D D I 株式会社）2012.08.23, 要約, 請求項 1（ファミリーなし）	1 - 1 4

☒ C 欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の 1 以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

0 2 . 1 2 . 2 0 1 3

国際調査報告の発送日

1 0 . 1 2 . 2 0 1 3

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号 1 0 0 - 8 9 1 5

東京都千代田区霞が関三丁目 4 番 3 号

特許庁審査官（権限のある職員）

戸島 弘詩

5 S

2 9 5 7

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 4 6

C（続き）． 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	竹之内隆夫, 複数のデータ提供のための匿名化, コンピュータセキュリティシンポジウム 2013 論文集, Vol. 2013, No. 4, 情報処理学会, 2013. 10. 14, 第 893-900 頁	1 - 14